CHAPTER I

INTRODUCTION

Digital video has taken the place of traditional analogue video in a wide range of applications due to its compatibility with other types of data (such as voice and text) [1]. The preliminary emphasis of research work regarding the compression is concerned only with minimizing the bandwidth. In the present multimedia world, the need for the storage and the transmission of the image and the video data is cumulative. This, in turn, leads to the improvement to be made in the compression methods. In addition to the real-time processing task of reducing the redundant video data, compression minimizes the bandwidth of the content and enables the fast transmission in the application-oriented environments [2]. The achievement of the compression efficiency without sacrificing the quality of the image and the content in the video is the most cautious one. The compression has been used in many of the applications. Some of those applications are cloud storage, transmission of medical and astronomical images and digital cinema [3].

1.1 IMAGE PROCESSING

Image processing is the primary entity considered in the video compression. This is because the video sequence to be compressed is processed as the image frames. In image compression, the problem is majorly concerned with the reduction of data required to represent the image. The redundant data reduction is the key for the image compression. Three basic types of redundancies in the image are coding redundancy, interpixel redundancy and psycho-visual redundancy [4, 5]. The coding redundancy in the image is because of the small length code words used. The interpixel redundancy in the image is because of the correlation between the image pixels. The psycho-visual redundancy is because of the ignorance of essential information by the human visual system. Image compression technology is named as the 'enabling technology' at present because of the vast area of the application hosted by the image compression. The applications include remote sensing, document and medical imaging, fax, space and hazardous waste management, control of remotely piloted vehicles in the military [5], etc. The redundancy data to be reduced from the image in the mathematical perspective is given as the amount of data utilized to transform an array of a 2-D pixel into the data set by means of statistics which are not correlated. The wavelet transform has been used as the efficacious tool for the image compression because of its energy compaction application.

The image represents the pictorial format of the data. The image has the 2-D representation of the data. The collection of pixels forms the image. The quality of the image increases with the increase in the number of the pixel. The image compression reduces the storage for the image by eliminating the redundant pixels from images. The image compression should be done without compromising the quality of the image. The image compression enhances the video compression process to be effective. The image compression with the effective algorithm completely removes the redundant pixels from the image. The image compression is completely different from reducing the size of the image. The image compression has many challenges.

The high correlation exists between the neighboring pixels of the image. Hence, the complete representation of the image cannot be done using the pixel values alone. The image representation with the correlated pixels has more redundant encoded information than the uncorrelated image. In this case, the image compression can be done only after the de-correlation of the pixels.

1.2 COMPRESSION SYSTEM

Image compression plays a significant role in the storage and transmission of image data. In addition to image data, audio data, voice data, animations, full motion video, and graphics are also sent. This is called multimedia data. Image and video data occupy a large portion of bandwidth during transmission. Thus, image data need to be compressed without or with small loss of quality. The major challenge is to develop efficient image compression algorithms.

The compression techniques can be broadly classified into two types. They are: lossy compression, and lossless compression.

In lossy compression, the information is lost during the decompression process. The lossy compression provides a high-quality compression through the loss of the contents. The results obtained during the decompression process differ from the actual content. The lossy compression provides better performance for the applications such as compression of the movies, sound files and a large quantity of the images. The loss in the information content does not affect the quality of the compression process until the loss is in very small amount. But, for the sensitive applications such as the compression of the computer programs, even the loss of a single bit of information makes the entire compression process to be unworthy.

The lossless compression performs the compression of the files without the loss of the data. But, this process has lesser quality than the lossy compression. For the sensitive applications such as the compression of the computer programs, even the loss of a single bit of information makes the entire compression process to be unworthy. These types of applications use the lossless compression. The lossless compression finds the application in the compression of the text files and the high-security video files. The compression ratio is used to identifies the compression quality and it is represented by the following equation,

$$\text{Compression ratio} = \frac{\text{Number of bits of the input image}}{\text{Number of bits of the compressedimage}}$$

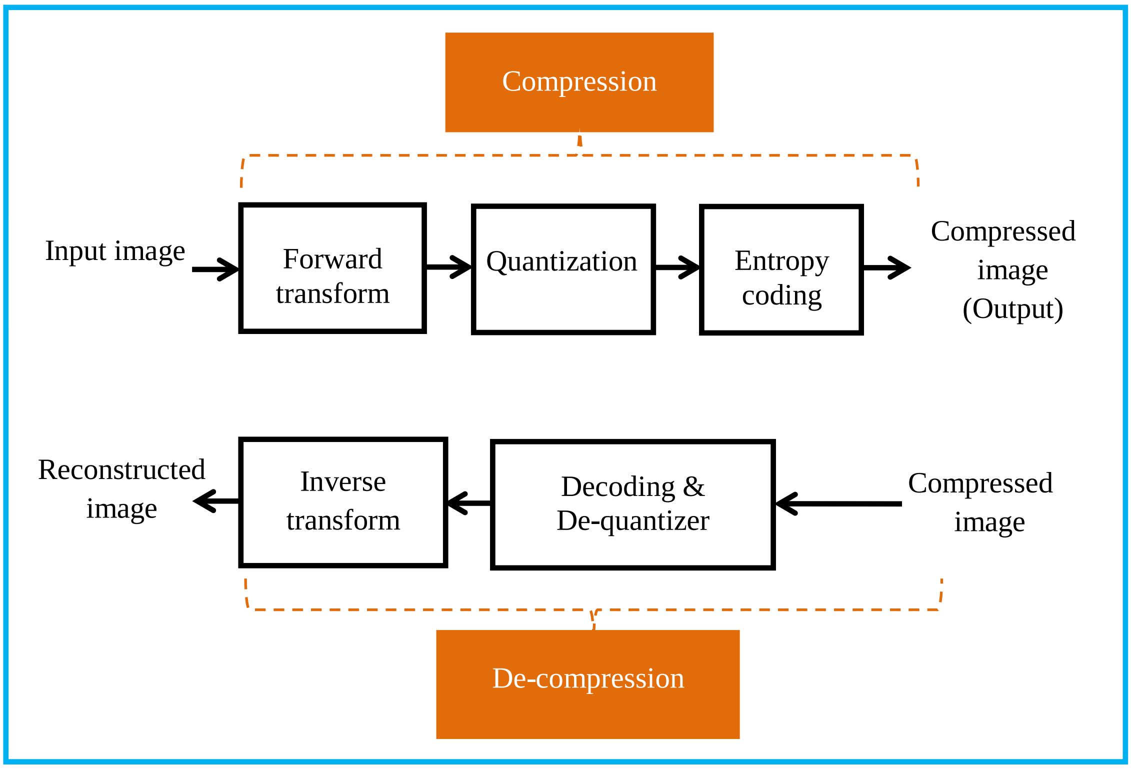

Figure 1.1: Basic Compression and De-compression System

The basic compression system [6] is shown in Figure 1.1. An image compression system contains three components: a forward transform, quantizer, and entropy coding.

Initially, the original image is transformed into the image of another domain, with high de-correlation concentrating more on the image information. The redundancy in the transformed images is removed and stored as a bit stream. The second block is the quantization. Based on some pre-established fidelity criteria, the quantization reduces the accuracy of the transformed output image. The final block is the entropy coding, in which the quantized image is coded into the fixed or variable length code words. The output of the entropy coding block is compresses image. The compressed image is used for the transmission with the low bandwidth in the application environments.

The de-compression is a reversible process. The compressed image is decompressed to reconstruct the original image. In the de-quantization stage, the accuracy based on the defined criterion is improved and results in the transformed image. In the Inverse transform stage, the transformed image is transformed from another domain into the original image domain. The reconstructed image may lose some information due to the procedure involved in the compression and results in the erroneous or distorted input image.

The performance of the compression system depends on the efficient transformation, quantization and the coding techniques used. The most powerful transformation technique used is the Discrete Cosine Transformation (DCT) [7] and the coding technique used is the Context Adaptive Variable Length Coding (CAVLC) [8].

1.3 VIDEO COMPRESSION

The video stream represents the continuous and rapid flow of the image files. The video combines both the image files and the sound files. The video compression performs the compression of both audio and the image information. The image compression should be done better because the video compression represents the number of reduced data needed to specify the images of the video. The video has the sequential arrangement of the images as the frames in the temporal and the spatial dimensions.

Once a video signal is digital, it requires a large amount of storage space and transmission bandwidth. To reduce the amount of data, several techniques are employed that compress the information without negatively affecting the quality of the image. The reduction in the data provided by the video compression over the intended video content pours the favor of minimum storage capacity and the bandwidth for transmission. The main objective of the video compression is safeguarding the quality of the data even after the redundant data reduction [9]. The efficient compression strategy reduces the overall size of the video without any diverse effect on the quality.

The basic video compression system comprises of both the compressor and the decompressor. The compressor used in the system acts as an encoder and the de-compressor acts as a decoder. The encoder in the system converts the raw video content (source video) to compressed video contents through elimination the redundant frames.

The compression efficiency of the video compression system depends on the functional components used for the removal of the redundancy [11]. The functional components in the video compression are shown in Figure 1.2. The redundancies in the video sequences are generalized into four types. They are: perceptual redundancy, spatial redundancy, temporal redundancy and statistical redundancy.

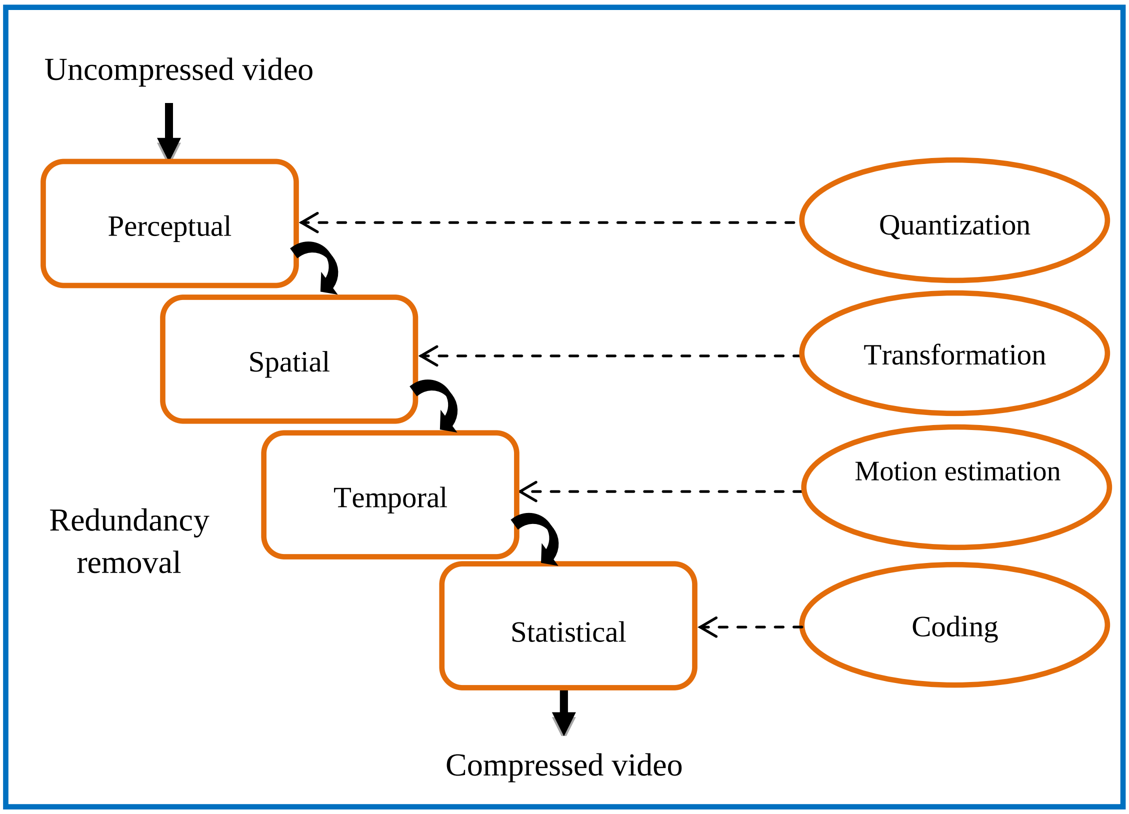

Figure 1.2: Functional Components in Video Compression

Perceptual redundancy (psycho visual redundancy) is associated redundancy in the frames which are not visible to the human eye system. The perceptual redundancy is eliminated by using the proper pre-filtering technique, quantization, chroma upsampling, object extraction, etc.

Spatial redundancy occurs in the neighboring pixels. The neighbouring pixel's value is identified by the correlations among the neighbouring pixels in the edges of the frames. The predicted pixel is subtracted from the original neighboring pixel, resulting in an error frame having lower entropy. These neighboring pixels are redundant pixels which can be reduced since they will not affect the quality of the image. The redundancy among the color components i.e. Red Green Blue (RGB) is called the spatial redundancy [10]. The exploitation of the spatial redundancy is achieved by the adaptation of the transformation. The integer transformation and intra-prediction are of the kinds used for the reduction of the spatial redundancies.

Temporal redundancy is the redundancy related with the inter frames. Temporal redundancy can be reduced using the motion estimation of any kind such as sub pixel motion estimation, motion block estimation.

Statistical redundancy occurs in the quantized data output which can be removed by the proper coding techniques used. Some of the coding techniques used for the statistical redundancy removal are huffman coding, adaptive variable length coding, arithmetic coding, etc.

1.4 IMPORTANCE OF VIDEO COMPRESSION

The details of video data such as, frame rate, duration, number of bits per pixel, transmission time needed for sending the video data and transmission bandwidth are clearly illustrated in Table 1.1.

Table 1.1: Video Data Transmission Bandwidth and Transmission Time

Duration	Bits per pixel	Uncompressed size	Transmission bandwidth	Transmission time	Transmission time after compression (42:1 ratio)
640×480, 1 min (30 frames/sec)	24 bits/pixel	1.66 Giga Bytes	221 Mbits/sec	Five days 8 hours	3.05 hours

For sending the uncompressed video sequence of 1.66GB, the transmission time required by the standard modem is five days 8 hours. For the application-oriented environment in the present world, it is waste one. The only solution is to compress the data before transmission and decompressed at the receiver side for the playback. For example, with a compression ratio of 42:1, the requirement of the video coding system can be reduced by the factor of 42 without any compromise in the video quality. The transmission time 3.05 hours required for after compressing video sequence with 42:1 ratio.

Video compressions include the merging of the unwanted information (redundancy) and reduce or remove it for the viewing of the video content. An effective video compression standard (format) provides the benefits such as no degradation in the visual experience of the video content.

1.5 VIDEO COMPRESSION STANDARDS

The video technology group associated with the generation of the video compression standards and evolution of the video standards over the time are discussed below.

1.5.1 Standard video technologies

There are two main standardized bodies for the video technology. They are

i. International Telecommunication Union – Telecommunication server (ITU-T)
ii. International Standard Organization (ISO)

The specific committee associated with the ITU-T for the multimedia data compression is called the Video Coding Experts Group (VCEG) [29, 30]. The committee responsible for the data compression in the ISO is called the Moving Picture Expert Groups (MPEG) [28]. The ITU-T coding standards are nothing but the recommendations, and they are denoted with the H.26x. The coding standards associated with the ISO are denoted with MPEG-x. In addition to the different moving picture and video coding expert groups, a new standard venture by joining hands of the ITU-T and ISO is developed which is called H.264 video standard.

The video compression standards associated to ITU-T and ISO are listed down. The advancement made in the consecutive video standards resulted in the compression with better visual quality and the compression efficiency.

A. H.120 Recommendation:

The first international digital video compression is the H.120 recommendation. H.120 was designed for natural continuous- tone visual content of any video or still picture. It was introduced in 1980 and got approved in 1984. The structure of the H.120 standard [29] contains coder with digital pulse code modulation, variable length coding, and scalar quantization. The bit rate control was provided by switching the motion compensation. The background prediction was not found to be optimal which was taken on focus for the future standardization.

B. H.261 Recommendation:

In 1990, H.261 [30] recommendation was adopted as an International standard by ITU-T. The designated purpose of the H.261 is the video conferencing applications. The bit rates provided by the H.261 are integer multiples of 65 kbps. The computation complexity of the H.261 standard is minimal, and the delay on the encoding of the bit streams is also minimal. The unique structure of the H.261 standard is the basis for the compression standard structure which is still found, including 8×8 block DCT, two-dimensional run level variable-length entropy coding properties, 16×16 macroblock motion compensation, and scalar quantization.

C. MPEG 1 Standard:

ISO released their first standard in 1992. The compression related with the CD-ROM was their main intention of MPEG 1. In addition to the video, the MPEG-1 layer 3 provided audio compression. MPEG-1 had a better visual quality of 1.5 Mbps and a bit rate range of about 1-2 Mbps [31]. The quality of MPEG 1 increases the bit rate. At higher bit rates, the performance was better than that of the H.261.

D. MPEG 2 Standard:

MPEG 2 [32] was the joint venture of the ITU-T and ISO and was developed around 1994. It is currently in widespread use. These techniques were similar to that of MPEG-1, but it can handle higher resolutions than MPEG-1. The encoding of the Digital Versatile Disk (DVD) is made by making use of MPEG-2. Bit rate range of the MPEG-2 was 2 to 10 Mbps. It makes the heart of broadcast-quality digital television for not only the Standard Definition Television (SDTV) but also for the High Definition Television (HDTV) with a step higher in popularity, picture quality, and bit rate. It efficiently handles the interlaced-scan pictures and hierarchical bit usage scalability.

E. H.263 Recommendation:

In 1995, ITU-T released H.263 [33] standard as a recommendation for the use of the video telephony. It is a video coding standard for video communication that has low bit rate over and mobile networks with transmission bit rates of around 10-24 kbps or above and Public Switched Telephone Network (PSTN). The rate, spatial and temporal scalability of the H.263 was similar to that of the MPEG-2. Its original target bit rate range was 10-30 Kbit/s, but this is extended to 10-2048 Kbit/s. The coding ability of the H.263 recommendation is same as H.261 and it uses the half or less than half the bit rate or even at very low bit rates. It provides the higher degree of performance for bit rates greater than 80 Kbps and it is superior to the H.261 standard.

F. H.263+ Recommendation:

H.263+ [34] as a series of H.263 was developed in1998. Even though H.263 was not recognized as MPEG, many applications indulging MPEG use H.263. Many digital cameras use H.263 for capturing video. The features of H.263+ in addition to the H.263 are compression efficient and tensile video formats and scalability. It was the first one to produce a greater degree of error resilience for wireless networks or the packet-based transport networks.

G. H.263 ++ Recommendation:

H.263++ recommendation [35, 36] was the series of H.263+. It was developed in 2000 by ITU-T. The bit rate of the H.63++ was more than that of H.263 +. The performance over the packet-based transport networks was further increased by the H.263++.

H. MPEG-4 Standard:

MPEG-4 [37] aims to improve the compression performance more than the MPEG-1 and 2. It supports a vast range of bit-rates, mainly focused on the low bit data rate. In low bit rate coding, MPEG-4 is similar to that of H.263. The basic concept held behind the MPEG-4 is object-based coding by which the scenes, background, and foreground can be coded independently with the prior video features and still-picture coding standards, namely zero-tree wavelet coding, hybrid coding of synthetic and natural video content, and segmented shape coding of objects [37].

I. H.264 Standard:

The ITU-T and ISO established a Joint Video Team (JVT) which is the combination of VCEG and MPEG groups. Based on the “back to basics” approach [39], they developed a new video compression standard in 2003. It is H.264 standard [40-42], which has also been incorporated into MPEG-4 under the name Advanced Video Coding (AVC). The primary intention of the JVT is to create a video standard with compression rate greater than the existing standards without affecting the visual quality. It provides the higher compression quality so that H.264/AVC [38] is quickly growing as the leading standard. This standard is adopted in most of the video coding applications, namely the Play station portable, the iPod and the TV broadcasting standards, namely Digital Multimedia Broadcasting (DMB) and Digital Video Broadcasting-Handheld (DVB-H). This is based on the block-oriented motion-compensation.

1.5.2 Overview of H.264/AVC

To provide better compression of JVT contains experts from both VCEG and MPEG developed H.264/AVC coding standard. The H.264/AVC provides the compression with the significant efficiency and simple syntax specifications. The H.264/AVC standard has 40% better gain than the existing standards. The H.264/AVC standard has a number of advanced features, such as I-frame encoding of video content, multi-frame inter prediction for removing temporal redundancy, de-blocking filters to eliminate the noise contents, small block-size transform coding, and Context-Adaptive Binary Arithmetic Coding (CABAC). These features in the H.264/AVC standard have a 40% higher bit rate saving than the other standards. The H.264/AVC model utilizes the Rate-Distortion (RD) optimization technique to compress the rate distortion occurring during the compression process. The H.264/AVC also provides smooth and continuous integration of the video coding in all the present protocols and multiplex architecture. H.264/AVC assists a number of applications, such as streaming of video, broad video casting, and video conferencing over wireless and fixed networks and various transport protocols [43].

A. Terminology:

The terminologies adapted in the H.264/AVC standard is described below,

Pictures: Coded video sequence in the H.264/AVC contains the series of coded pictures. This is called pictures.

Frame or fields: The coded picture is represented as a single field or the whole frame [41]. Interlaced and progressive [40] are the video types supported by the H.264/AVC. A frame in the interlaced video is decomposed into two fields constituting an odd number of the picture line and even number of the picture line, but in the progressive video, frames are not divided. The H.264/AVC contains three types of frames; They are I-frame, P-frame, and the B-frame. I-frame of the video can be predicted by using the Intra prediction. The P-frame is the Predicted frame makes use of both the intra-prediction and the Motion estimation. The B-frame is the Bi-directionally predicted frame. It depends on both the predicted frames and the future incoming frames.

YCbCr Color space: The color space utilized by the H.264/AVC divides a color representation into three components, namely luminance (Y) which represents brightness and chroma components Cb and Cr.

The two chroma components Cb and Cr represent the extent to which the color deviates from gray toward blue and red, respectively. In H.264/AVC as in prior standards, a YCbCr color space is used to reduce the sampling resolution of the Cb and Cr chroma information.

Macroblock: The coded pictures are subdivided into smaller units called macroblocks. Coding schemes involved in the H.264/AVC are block based.

Slices: A video picture is coded as one or more slices consist of an integral number of the macro blocks in it. The coded slices are of different types.
They are: i) Intra (I) slice, ii) Predicted (P) slice, iii) Bi-Predictive (B) Slice, iv) Switching 'P' slice and v) Switching 'I' slice.

B. Profiles:

The subsets of the coding tools intended for different classes of the applications are called profiles [44]. There are about three files. They are Baseline Profile (BP), Main Profile (MP), and Extended Profile (EP). BP profile is applicable for the real-time casual services, like video conferencing and video phone. MP is applicable to the television broadcasting and digital storage media. EP is applicable for the multimedia services over the internet. In addition to the intended class application profiles, extended fidelity range profiles are also present in the H.264 [32] for applications, like content contribution, post-processing, and studio editing.

The size reduction rate of the digital video file by the H.264 standard is 80% more than motion joint expert group format and 50% more than MPEG-4 standard, without any degradation in the visual quality. In addition to the reduction in the overall size, the transmission bandwidth, and the storage space requirement are also reduced by the application of the H.264/AVC standard. Irrespective of the bit rate, better visual quality attainment is the precise entity of the H.264/AVC (Advance Video Coding) standard is also known as industry video coding standard [39] because of the broad range of application provided by the industrial sector from mobile video applications that have low bit rate to HDTV. The advanced coding tools in the H.264/AVC coding standard are spatial-domain intra-prediction, improved in-loop deblocking filter, multiple reference frames, and variable block-size motion estimation, an MV with quarter pixel accuracy, Rate-Distortion (R-D) optimization, and entropy coding.

The coding performance increases regarding three measures over the existing video standards. They are Peak Signal to noise ratio, computational time, and Bits per pixel. The motion estimation of the H.264/AVC can be improved by increasing the prediction rate of the motion vectors and also by utilizing fast search algorithms [23-26]. The joint venture assigns the name of MPEG/AVC to the H.264/AVC video coding standard, representing modern in the video compression. The advantage of the H.264/AVC originates because of the property that in addition to the normalization of the decoding process, the encoding process is also normalized in the MPEG/AVC. In video coding, the visual quality and the compression efficiency are dependent on the motion estimation. The motion estimation process searches the specific area within the frame that causes changes when compared with the previous frame. In motion estimation process, the video frames are first divided into macro-blocks. The generic hybrid coding framework is applied to the efficient coding and reckoning of residual frames with the reference frame in which the macro-blocks are displaced [17].

The complexity reduces with the minimization of the factor needed for selecting the candidate motion vector. The candidate motion vector checking in the matching algorithm varies with the matching strategies and the concerned search pattern, i.e., cross, diamond, octagon, triangle, hexagon, etc. Different shapes and sizes of search pattern have a huge influence on both the speed and accuracy. The performance of motion estimation is based on three factors: motion model, the block size & shape, and matching the precision [19]. The prediction of the residual frames in the H.264/AVC standard is achieved by the transform coding same as that of MPEG-2. But the variation is in the type of the transform coding used. Instead of 8×8discrete cosine transform, H.264/AVC standard utilizes the 4×4 or 8×8 integer transform. In the integer transformation, the transformation operation involves only the addition and shift operations and the mismatch is never happened among the forward and inverse transforms same as that of the DCT. This reduces the computational complexity concerned with the transform coding in the H.264. If smaller size transform coding is used, i.e. 4×4, the block noise reduction achieved in the decoded video is significant. In H.264 encoding, the transform size and the block size for motion compensation or intra-prediction are independent to each other [38].

Figure 1.3 depicts the overall block diagram of the H.264/AVC codec. Input video sequence is converted into the image frames, and the images frames are again sliced into the macro blocks. The major processes in the AVC standard are the motion estimation and the motion compensation. The video compression is used to eliminate the redundancies presented in the video frame. The frames in the video may be an inter frame or an intra-frame. There are two modes of coding in the AVC standard; they are intra-mode and inter mode. The current block is compared with the candidate block for generating the predicted block.

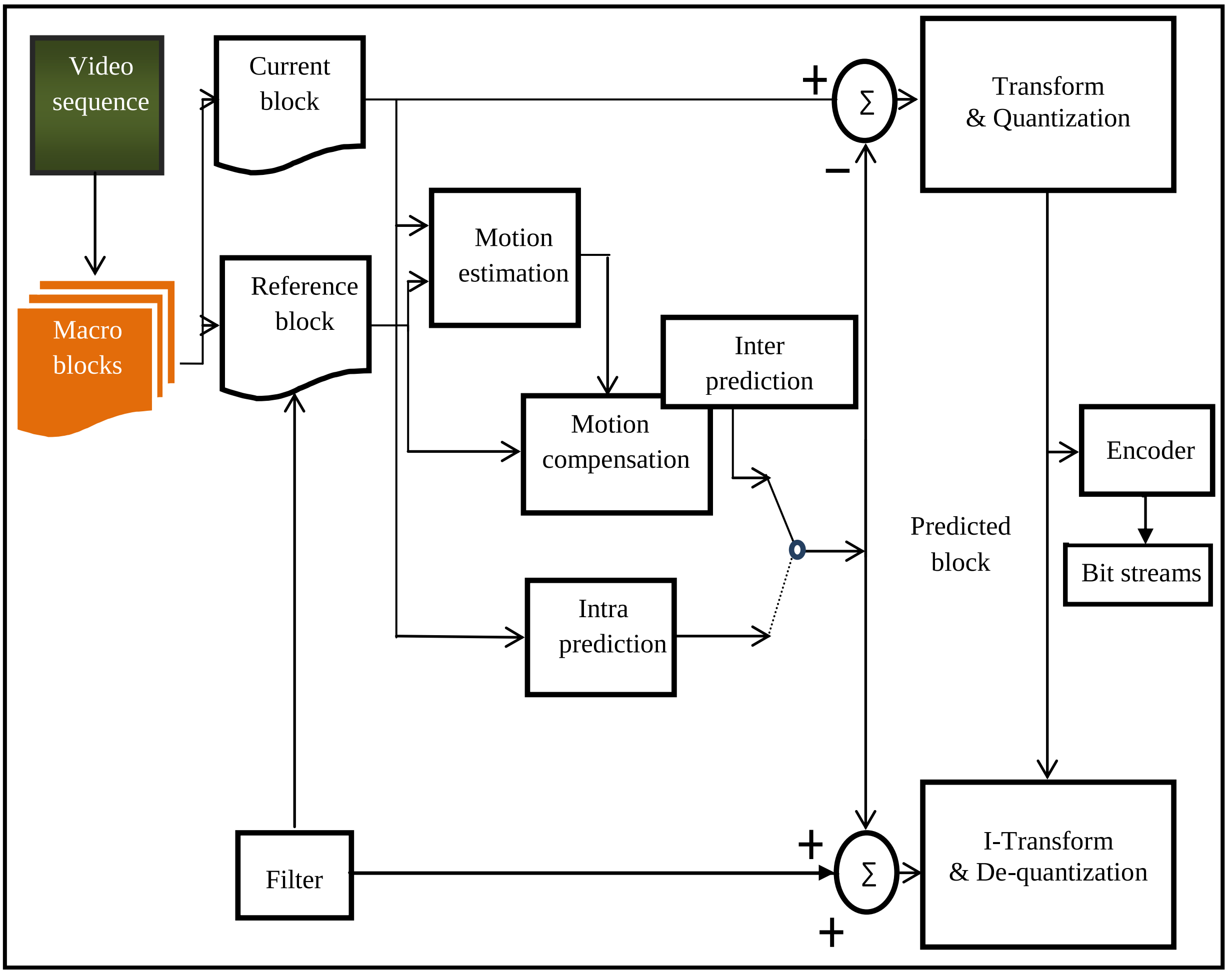

Figure 1.3: Block Diagram of H.264/AVC Standard

The motion compensation prediction is exploited for the inter mode with the help of the de-blocking filter. Intra prediction is achieved using the reconstructed block. A residual block is produced by subtracting the predicted block from the current block. The residual data in the block of the considered frame is transformed into another domain using the discrete cosine transform and quantized eliminating unnecessary coefficients. The quantized coefficients are encoded into bit streams and transmitted.

The deviation in the H.264/AVC coding structure and the other coding is because of the intra-frame prediction and the de-blocking filter.

The H.264/AVC standard utilizes the DCT for transforming the video contents into the frequency domain. This process improves the compression of the major blocks with less loss in the data. The main aim of the transformation is to separate the independent frames in the source video content.

The encoded information from the encoder is quantized with the use of the quantization block. The quantization process makes the video coding process to be lossy, because the quantization results in the loss of the frames. This is due to the rounding errors and the quantization noise. The data lost in the quantization process is in small quantity and can be ignored during the compression process.

The inter prediction block allows the prediction of the major blocks from the other major blocks. The H.264/AVC has the major blocks of the size 16×16 to the 4×4 with the motion vectors. The motion vectors allow the positioning of the pixels in the image frame of the video content. The inter-frame prediction predicts the image frame with the temporal redundancy. It exploits the prediction by providing the weights to the various redundant frames. The major block prediction is done by assigning multiple reference frames to the major block. The best matching algorithm predicts the major block of the frames.

The first frame in the major block of the video content has the less similarity with the previous frames in the major block. Hence, the I-frames cannot be predicted from the previous frames in the major block. I-frame prediction is done through the Intra-frame prediction since the frame depends on itself. In intra-frame prediction, the spatial redundancies associated with the video are utilized. The intra-encoded frame is of larger size because of a huge volume of the information is presented in the frame without temporal information. In the intra-coding process, the efficiency is increased by considering the spatial correlation between the motion blocks [13].

1.6 TYPES OF VIDEO COMPRESSION

In image compression, the image is represented with a minimum number of bits. The video compression is of two types. They are lossy compression [12] and lossless compression [13]. The video compression reduces the storage space for the video contents.

The video compression can be done in two ways. They are,

i. Lossy video compression
ii. Lossless video compression

i. Lossy video compression:

This method compresses the video content in an effective way but, there is a loss of the video content during the decompression process. The lossy video compression finds its applications in the storage of the movie files, video calling and video conferencing. During the process of decompression, it is not possible to retrieve the compressed data in lossy compression. The image quality degradation by the lossy compression is done to meet the target data rate suitable for the storage and the transmission in the envisioned source. Application of the lossy compression is a transfer of the image and video data through the band limited network. The major issue regarding the lossy compression is that how much quality of the image can be degraded for the given data rate.

ii. Lossless video compression:

This method compresses the video contents with a low quality but, has no loss in the video content during the decompression process. The lossless video compression is used commonly in the highly sensitive applications such as live streaming of the video, medical imaging, and military applications. The lossless video compression is preferred over the lossy video compression since it does not lose any data. The lossless video compression uses the DCT and wavelet transforms to perform the video compression. In lossless compression, the image retrieved from the decompressed image is same as the original image. Graphics Interchange Format (GIF) video is an example of the lossless compression. The lossless compression cannot be used in the video surveillance applications.

The compression approaches for the lossy and lossless compression are distinctive. The major approach enrolled in the lossless compression is the predictive coding or the entropy coding [14]. In predictive coding, differential pulse code modulation is used, and the prediction is obtained by the linear combination of the previously decoded neighbors. Run length encoding, Huffman coding or Arithmetic coding is used for the Entropy coding approach. In certain cases, the context modeling [15] is included in the entropy encoding, for the determination of the probability distribution of the symbols involved in the context.

In lossy compression, approaches like transform coding are used. In transform coding [16, 17], correlations between the pixels are reduced by the transformation itself. The transformed coefficients are truncated into the bit streams, and the coefficients are reorganized based on the properties, providing scalability to the compressed bit streams.

The predictable properties along with the video compression techniques eliminate redundant video frames to perform the compression. The lossless compression is preferred over lossy compression for compression of the video streams. The two predictable properties used in the video compression are,

i. Intra-frame prediction
ii. Inter-frame prediction

The intra-frame prediction predicts the spatial redundancies in the image frames of the video content. The inter-frame prediction predicts the temporal redundancies occurring between the subsequent frames of the video.

The prediction of data is attained through: with inter-frame prediction and intra-frame prediction [18, 19]. By minimizing the temporal redundancy within the consecutive frames before intra-frame coding, the inter-frame prediction is achieved. Block motion estimation [19, 20] is the commonly used technique for the inter-frame prediction. In the video contents like a live news program, the frame to frame scene change is very minimal. By taking advantage of this property, the compression is managed by only storing the difference instead of storing the entire frame. The motion estimation and compensation [18, 19] is the technique relieved with the advantage of storing the difference between the frames. Motion estimation is the method in which the current frame's object is located from a candidate frame. There are three frames are involved in the motion estimation and compensation process, namely I-frame, P-frame, and B-frame. The key frame which is not compressed in a designated video is called as the Intra-frame or I-frame.

The frame is stored as non-identical to that of the previous reference frames is called as the Predicted frame (P-frame). P-frame is the term used to define the forward predicted pictures. The prediction is made from an earlier picture, mainly an I-frame, so that require less coding data (≈50% when compared to I-frame size). The amount of data needed for doing this prediction consist of motion vectors and transform coefficients describing prediction correction. It involves the use of motion compensation.

B-frame is the bi-directionally predicted frame. B-frame is created based on not only the previous frames but also the future frames. This kind of prediction method occupies less coding data than P-frames (≈25% when compared to I-frame size) because they can be predicted or interpolated from an earlier and/or later frame. Similar to P-frames, B-frames are expressed as motion vectors and transform coefficients. In order to avoid a growing propagation error, B-frames are not used as a reference to make further predictions in most encoding standards. However, in newer encoding methods (such as H.264/AVC), B-frames may be used as reference.

The typical Group of Pictures (GOP) structure IBBPBBP is shown in Figure 1.4. The I-frame is used to predict the first P-frame and these two frames are also used to predict the first and the second B-frames. The second P-frame is predicted also using the first I-frame. Both the P-frames join together, to predict the third and fourth B-frames.

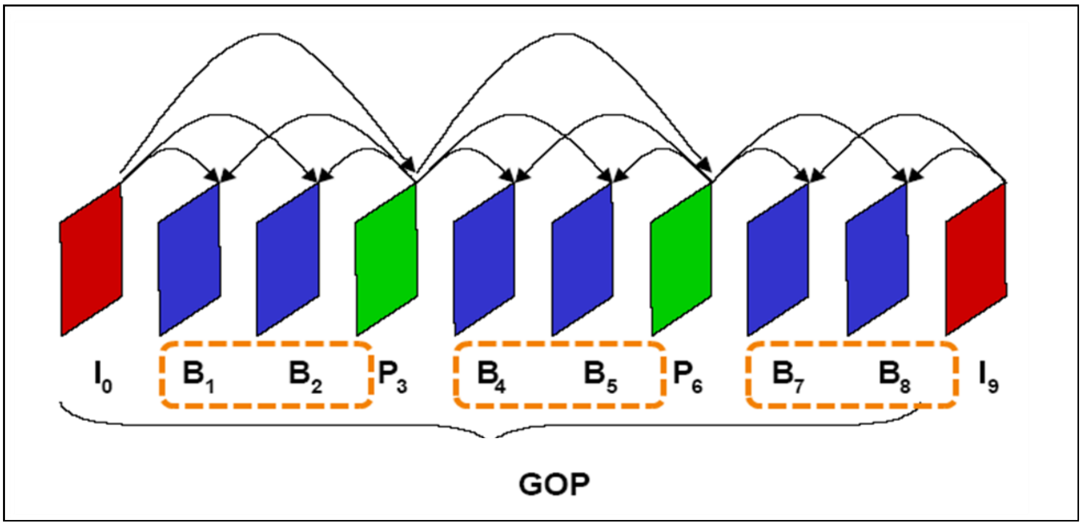

Figure 1.4: Group of Pictures

The encoding order and playback order of the GOP is shown in below.

Encoding order: I0, P3, B1, B2, P6, B4, B5, I9, B7, B8.

Playback order: I0, B1, B2, P3, B4, B5, P6, B7, B8, I9.

The basic assumptions considered for the motion estimation are,

i) Illumination is constant along the motion path, and
ii) The blocking problem is not present.

In the first assumption, the changes in illumination over the time period are neglected which is not necessary for the motion. The second assumption ignores the problem of the scene change between the frames and the uncovered background. The Block Motion Estimation (BME) [20] divides the frame into a number of blocks and each block of the current frame is estimated using the block in the reference frame which size is equal to the size of the block in the current frame. The relative distances between a reference block and the current block is called as the Motion Vector (MV). The prediction error comprises of both the reference and current block is converted into a coded form and transferred alongside the motion vector. The redundancy between the nearest motion vectors is made available to use by encoding the motion vectors among the current frame and the previous frame. The compression efficiency concerned with the block motion estimation can be reduced by coding each frame independently with the image coding techniques [21]. The motion estimation methods in the video compression are of different types [22].

i. Block-based motion estimation
ii. Variable block size motion estimation
iii. Sub-pixel motion estimation

i. Block-based motion estimation:

Block-based motion estimation is widespread among the varying standards of the compression. The block-based motion estimation divides the frame into blocks in which the size of the block is square. Each block in the frame is matched to the reference frame depends on the block distortion measure. The perfect differences among the current block and the reference block are summed to form the popular BDM. The displacement resulted in the matching of the motion vector. The search over the entire frame is confined by a search window.

ii. Variable block size motion estimation:

Variable block size motion estimation is similar to that of the block-based motion estimation. In variable block size, the blocks associated with the frame are partitioned into several blocks of variable size. Consider an example, 1 MB of size 16×16 video sequence is decomposed into two rectangular blocks 8×16 or 16×8, or decomposed into four square blocks of 8×8. If four 8×8 blocks are used as the macroblocks, it is again spliced into sub-macro blocks. The block with the reduced size is used for the motion vector estimation which allows more accurate prediction in the motion compensation process.

iii. Sub-pixel motion estimation:

The sub-pixel motion estimation scheme can provide significantly better compression performance than other types. But the computational complexity concerned to the pixel level is vast. The coding efficiency of use of the sub-pixel motion estimation increases the bit rates and the resolution of the video in the coding system.

1.7 MAJOR CHALLENGES

Due to the significant applications of H.264/AVC for video compression, researchers have tried different variants of video compression model based on this standard. The most of the research is on the motion estimation to find the motion vectors through different search algorithms. The search algorithm requires more computational time to find the best motional block.

i. The main aim of the searching algorithm in the video coding process is to find the best matching block of the current frame based on the reference frame. The best matching block should satisfy the two constraints namely, rate & distortion, defined in the trade-off criterion.
ii. As the most of the videos are symmetrical in nature, the square search finds difficulty in preserving the motion discontinuities like, zooming, rotation, fast moving objects and so on.

To accomplish these discontinuities, hexagonal search, and the octagonal search was introduced to easily obtain the best matching block by covering the search area's edge points. The major challenges in the motion estimation are computational complexity, the search pattern's size and shape.

i. To determine the best matching point, it requires more search points, then the complexity is high, with the increasing the search points will further enables the gain in video compression.
ii. Finding of a better matching point to estimate motion field is a judicious progression in video coding process.
iii. To designing of rate-distortion and trade-off played a major role in selecting the blocks for further encoding.

1.8 PROBLEM STATEMENT

The search algorithms of different kinds are used in the [4, 12, 77-79, 91] for the motion estimation and compensation. In some motion estimation algorithms, the refining of the search classification affects the motions vector vigorously [95]. The search pattern that can handle the tricky motion such as the zooming, rotation and fast moving objects must be chosen for the motion estimation. The square search found difficult to preserve the motion discontinuities.

i. In order to cover the motion discontinuity, increased number of searching points, compression performance problem, the Adaptive Order Square Hexagon (AOSH) search algorithm was developed on H.26/AVC.

ii. The search algorithm partitions the areas based on the direction of the search. With the unknown direction of the search, the searching points utilized for the selection of the proper direction of the search is increased. In every possible direction, the numbers of search points get increased because of the uncertainty. To find the motion discontinuity, compression performance and computation time, the Adaptive Order Cross Square Hexagon (AOCSH) search algorithm was developed.

iii. Context-adaptive binary arithmetic coding is employed to find best compression for the video data.

The main purpose of this research work is to propose motion estimation algorithms for H.264/AVC standard of video compression using modified adaptive order search methods to reduce the computational time and compression efficiency of the H.264/AVC encoder, at the same time without degrading the picture quality as compared to the existing methods.

1.9 OBJECTIVES OF THE RESEARCH WORK

The objectives of the research work are formulated as follows:

i. To study and analyze the various motion estimation algorithms.

ii. To identify the various challenges involved in the motion estimation algorithms for obtaining the better performance in H.264/AVC.

iii. To design and develop a new search algorithm for motion estimation in H.264/AVC to improve the compression efficiency and to reduce encoding time.

iv. To analyze the performance of proposed algorithm and compare with existing algorithms for various videos in terms of PSNR, SSIM, and computational time.

1.10 CONTRIBUTIONS OF THE RESEARCH WORK

The main contributions of the research work are given as follows:

An Adaptive Order Square Hexagon (AOSH) search algorithm is developed for the estimating the motion and the weighted tangent trade-off function for the evaluation of the best matching block. The proposed AOSH algorithm minimizes the number of the search points.

An Adaptive Order Cross Square Hexagon (AOCSH) search is utilized for the motion estimation searching process, and the fuzzy tangent weighted trade-off function is proposed for the evaluation of the best matching blocks. Primarily, using the cross search, the direction of the search is identified, thereby minimizing the number of search points.

1.11 ORGANIZATION OF THE THESIS

The thesis is organized as follows:

In Chapter 1, introduces the video compression, and overview of H.264/AVC. From this discussion the challenges of video compression, objectives of research is defined.

In Chapter 2, informs literature survey on motion estimation and search algorithms that are used in video compression applications is presented.

In Chapter 3, discusses about motion estimation for H.264/AVC, Elastic motion estimation method and its impact on the video coding is discussed.

Chapter 4 proposes an adaptive order square search, adaptive order hexagonal search and development of AOSH search algorithms with tangent weighted trade-off function for the motion estimation with detailed explanation.

In Chapter 5, proposes a development of AOCSH search algorithm, and fuzzy tangent weighted trade-off function for H.264/AVC standard of video compression.

Chapter 6 discusses about AOSH algorithm simulation is performed with different videos and the performance evaluation is compared with the existing algorithms H.264/AVC and elastic motion estimation method and also the performance analysis of AOCSH with fuzzy tangent weighted trade-off criterion is performed using SSIM, PSNR and computation time metrics.

Lastly, conclusions and future research directions of this research work are given in Chapter 7.