A Framework of Video Compression and Decompression

Neetish Kumar[1*], Deepa Raj[2]

[1]*Research Scholar, Department of Computer Science, BBAU(A Central Univ.), Lucknow,U.P , India. E-mail:neetish08537@gmail.com.*
[2]*Assistant Professor, Department of Computer Science, BBAU(A Central Univ.), Lucknow,U.P , India. E-mail:deepa_raj200@yahoo.co.in*

Abstract

As Multimedia and communication technology is on peak nowadays and keep growing day by day, the vast and voluminous content present on the web is video data. It needs to be compressed as it is of draconian in nature if not compressed. A lot of compression techniques are developed for handling the video data. If it is keenly observed then it is found that search techniques play a pivotal role in video compression. In Video compression, apart from the search techniques, some redundancies should be exploited. Various types of redundancies are the main contents of an uncompressed video. Spatial, temporal, Statistical redundancies occur based on their nature and properties and exploited accordingly without any loss of actual information in video. Block based motion estimation serves as the backbone of video compression and the search techniques used in the block based method play a pivotal role in detecting the motion vector. The encoding of motion vector and residual results in a compressed video form. This paper elaborates on a new kind of framework, having a slightly different type of technical strategy of video compression. All the search techniques used in this paper are pentagon search for block matching motion estimation.

Keywords: *Motion estimation, Motion compensation, Encoding, Decoding, Motion vector.*

1. Introduction

- **Introduction about the video:** Before going in depth of motion estimation and motion compensation, it must be understood the structure of a video properly. A normal video is composed of different types of shots. A shot mostly contains GOPs (group of frames) and those GOPs are core terms for compression because it contains different types of frames namely I, P and B frames. For analysis purpose a data structure is used for traversal of macro block, and then it comes to a matter of blocks and pixels that is the basic thing for a video. A semantic diagram of the normal video is shown in Figure 1.

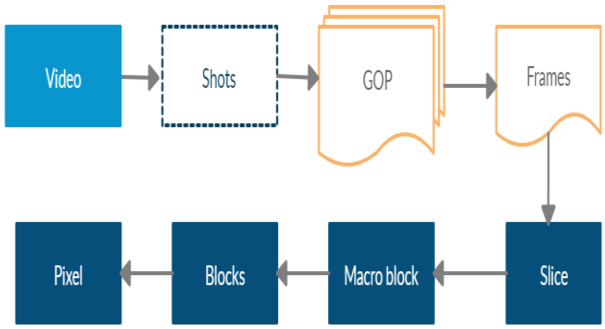

Figure 1. Organisation of video data

ISSN: 2005-4262 IJGDC

- **Motion estimation and motion compensation:** As motion estimation is the crucial part of video compression, a brief mechanism of it must be described. Firstly, the $\boldsymbol{Image_t(x,y)}$ is split in block $\boldsymbol{Blocks_t[n](x',y')}$. The function of motion compensation is to generate $\boldsymbol{Blocks_t[n](x',y')}$ from any part of $\boldsymbol{Image_{t-1}(x,y)}$. So, we have to find out the best possible matching block $\boldsymbol{Blocks_{t-1}[n](x'+mx,y'+my)}$ not strictly nearby of 16x16 boundary. Here, ***mx,my*** is called motion vectors. Basically, the encoder transmits the error and the motion vector for each block. The error can be calculated between the target and reference as $\boldsymbol{Err_t[n](x,y)=Blocks_t[n](x',y')-Blocks_{t-1}[n](x'+mx,y'+my)}$. The processor estimating *mx,my[n]* for every n such that $\boldsymbol{Err_t[n](x,y)}$ is minimized is called Motion estimation. Whereas, the process of constructing $\boldsymbol{Blocks_t[n](x',y')}$ from It image pixels and ***(mx,my)[n]*** is called Motion compensation. The error image is what gets transmitted. Finally, decoder can redo motion compensation on its own using motion vectors and the error image to make the final re-construction of the image. The pentagon search is used for searching the motion vector throughout the paper for encoding and analysis. Search pattern of pentagon is shown below in Figure 2 [12].

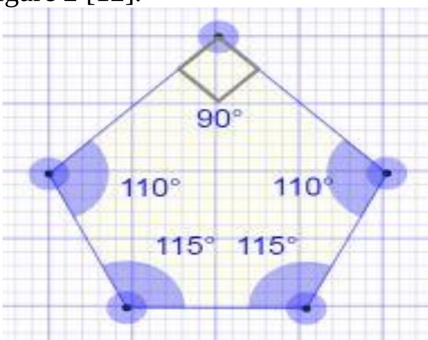

Figure 2. Shape of the pentagon

- **Compression strategy of videos**

  Steps in video compression

  - Motion estimation & Motion compensation: These two crucial parts are discussed in above section.
  - Transform and quantization: The basic transform coding process is almost similar to that of every codec mechanism. The process includes a forward transform and quantization followed by zigzag ordering and entropy coding.
  - Inverse transform: It is used in the reconstruction process. The reconstruction process includes an inverse quantization and inverse transform followed by motion compensation.
  - **Some codecs and their working principles**

There is some limitation on the certain simulator to analyze the result of video frames. In general these frames are of big size which takes too much time for analysis. As suggested in various literature, all the simulation for video compression is performed on QCIF video format due to prevent the heavy computational complexity. Time taken for the simulation of normal/ HD video takes through MATLAB usually takes much more time. Hence for analysis purpose CIF or QCIF are used. Spatial resolution and other properties for different formats are given in Table 1.

Table 1. Different formats of videos and its properties

Format	**Spatial Resolution**	**Frames/Second**	**Uncompressed Data Rate (bits/s)**
QCIF	176 × 144	15	4 561 920
CIF	352 × 240	30	30 412 800
ITU-R601	720 × 480	30	124 416 000

HDTV	1280 × 720	30	331 776 000
HDTV	1920 × 1080	30	746 496 000

MPEG 1:
Moving Picture Experts Group (MPEG) committee introduced the MPEG-1 as the first public standard for video compression. It was developed by ISO and was capable of compressing a video at the compression ratio of 26:1 and 6:1 for video and audio compression without excessive loss of quality. The standard consists of the following five parts.

MPEG 2:
The MPEG-2 Codec was developed in 1995 keeping, in mind to extend the compression strategy of MPEG-1. A step forward it not only compresses a video but also covers the scope of bandwidth usage. It is suitably designed for digital television broadcasting applications. It is capable of handling the bit rate of TV broadcasting that lies between the range of 4 and 15 Mbps (up to 100 Mbps), such as Digital High Definition TV (HDTV), Interactive Storage Media (ISM) and cable TV (CATV). Profiles and levels were introduced in MPEG-2.

MPEG 4:
MPEG-4 is a method of defining compression of audio and visual digital data. MPEG-4 is still an evolving standard and is divided into a number of parts for its simplicity and betterment of this codecs. Its supplimentry parts are MPEG-4 Part 2, MPEG-4 part 10 (MPEG-4 AVC/H.264 or Advanced Video Coding) etc. It was developed in October 1998 and it empowers audio as well as video in low bit-rate networks and allows the user to interact with the objects. Video object coding is one of the most prominennt features originated in MPEG-4 that enables arbitrary shape in spite of a rectangular frame for block matching. This type of compression incorporates the inclusion of shape, motion, and texture.

Table 2. Shows the different compression techniques of MPEG family

S.No	Year	Standard	Publisher	Popular Implementation	Bit Rate
1	1992	MPEG-1	ISO	Video-CD	1.5 Mb/s
2	1994	MPEG-2	ISO,ITU-T	DVD video, digital video broadcasting	>2Mb/s
3	1998	MPEG-4	ISO	Video in internet, DivX	Variable

H.261:
International Telecommunication Union (ITU) developed this standard in 1990. Motion compensated temporal prediction are easily handled by it. Due to the limitation of hadling voluminous visual information at that time, it supports two resolutions, namely, Common Interface Format (CIF) with a frame size of 352x288 and quarter CIF (QCIF) with a frame size of 172x144. It provides the coding and decoding services at the rate of p x 64 kbit/s, where p is in the range 1 to 30. As like of MPEG standard it uses discrete cosine transform(DCT) and performs both INTRA as well as ITNER frame encoding. The prominent features of H.261 is prediction, block transformation (spatial to frequency domain translation), quantization and entropy coding without degrading the video quality. Basically loop filtering serves as the backbone for improvement in video quality with extra processing power. It is mostly used in video conferencing and video communications systems. Its usage includes studio based video conferencing, desktop video conferencing, surveillance, monitoring, computer training, and telemedicine.

H.263:
With some limitations over the video conferencing application in H.261, H.263 is designed and developed by the ITU-T Video Coding Experts Group (VCEG). A step forward H.263 standard not only supports video conferencing but video telecommunication too. It was approved in early 1996. the compression core of the MPEG-4 standard lies in H.263. Variable block size compensation, overlapped block motion compensation are the prominent ideas of H.263 standard. While H.261 allows 18- 24

ISSN: 2005-4262 IJGDC

kbps for moderate quality over a regular phone line or wireless medium, H.263 allows 28-33 kbps for the high quality. It is used for various applications on the internet eg. Flash Video Content. It supports five resolutions: QCIF, CIF, SQCIF, 4CIF, and 16 CIF.

H.264:

H.264 or MPEG-4 Part 10, Advanced Video Coding (MPEG-4 AVC) was developed by the ITU-T Video Coding Experts Group (VCEG) together with ISO/IEC. The developer of MPEG series and H.26X series came together for this project and designed the best Codec. H.264 encoder provides better improvement in compression efficiency than the previous standards. It is basically a block-oriented motion-compensation-based video compression standard. H.264 is mostly used for encoding the content of Blue ray disc, pervelent in video conferencing, television broadcasting etc.

Table 3. H.26x family series

S.No	Year	Standard	Publisher	Popular Implementation	Bit Rate
1	1990	H.261	ITU-T	Video Conferencing, Video Telephony	P*64 kb/s
2	1995	H.263	ITU-T	Video Conferencing, Video Telephony, Video on mobile phones	<33.6 kb/s
3	2003	H.264	ISO,ITU-T	Blu-ray, DVD, Digital video broadcasting, HDTV, Apple TV	10’s to 100’s kb/s

2. Literature Survey

P. B. Penafiel and N. M. Namazi introduced a new framework for video compression that considers noise applicabitily in frame sequence directly in order to find the optimal compression ratio. By exploiting the temporal and spatial redundancies they found the compressed video. Processing is performed in blocks of N frames stored in a video buffer. Encoder and decoder are synchronized before the transmission of a new block [1]. K. Minoo and T. Nguyen suggested a framework for efficient entropy coding of data through parametric distribution model. Maximum a posteriori (MAP) or Maximum Likelihood (ML) and other statistical parameter estimation techniques are used in their modified framework that basically applies on entropy coding [2]. L. Xiaoli et. Al. Demonstrated a framework based on perspective to improve the degree of parallelism of the on-chip peripherals and the core for video compression [3]. M. Mody suggested a framework that doesn’t need much more hardware support for video compression. The proposed model, VCF also defines simple primitive (e.g. threads, queues, semaphores, messaging, scheduling scheme etc) to enable it without using any hardware support [4]. S. Moni and S. Sista elaborated a framework for Images and video compression that are being extensively used in numerous areas such as video-conferencing, multimedia documentation, telemedicine, high definition television (HDTV) etc [5]. Xuguang Yang and K. Ramchandran discussed a novel region-based video compression framework based on morphology to efficiently capture motion correspondences between consecutive frames in an image sequence. They formed “clusters” based on thier idea of motion field associated with typical image sequences. These segmentation leads to the distinct objects or regions in the scene, that can be efficiently captured using morphological operators in a "backward" framework that avoids the need to send region shape information [6]. L. Herranz and J. M. Martínez worked on video summarization that provides compact representations of video sequences. It is basically based on the visualization, rather than the information conveyed [7]. J. Nightingale et.al. proposed an idea that focuses on improvements to video compression efficiency of

HEVC. In their work, they considerd all the barriers towards the development of HEVC framework. This study filled the gap at every stage of visual application, testing & evaluation, video streaming and packet loss in HEVC [8]. D. Schonberg et.al. presented a framework for compressing encrypted media, such as images and videos. They first develop statistical models for images before extending it to videos [9]. F. Kamisli demonstrated the different idea of video compression in which when spatial prediction is used, temporal information is ignored [10]. C. Zhang et.al. proposed a novel algorithm called spatially varying transform (SVT) is proposed to improve the coding efficiency of video coders. SVT enables video coders to vary the position of the transform block, unlike state-of-art video codecs where the position of the transform block is fixed [11].

3. Proposed framework for video Compression & Decompression

Till now there are a lot of video codecs developed, our proposed framework is a step forward in this regard. Basically at the back end i.e search strategy, a new pentagon search method is used that searches the true motion vector so much efficiently that further encoding method gets benefitted by it. Before knowing the complex process of compression it must be investigated first that what a video frame consists of. Basically video is a sequence of kinetic frames. These frames are bundled together by its properties, these bundles are known as GOP. In a GOP there are three types of frames I, P and B frames. All three types of frames are encoded in different ways. The first frame of the GOP is always an I frame that is encoded just like the still image. Figure 2 represents the diagramatic representation of GOP that deals with the coding of different frames.

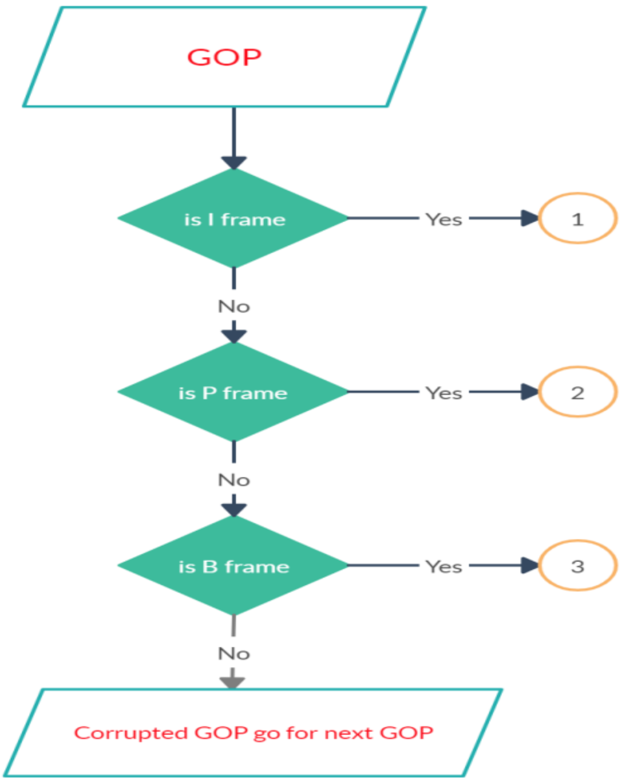

Figure 3. Encoding of different frame through GOP

3.1 Encoding of different Frames

a) Encoding of I frame: I frame is encoded just like the JPEG image. The first frame of every GOP is I frame. It is also known as INTRA frame as it has only the spatial redundancy. This is the only frame that is intracoded and also known as error-free frame because at this stage no prediction is needed for the encoding. It acts as a reference frame for the upcoming P and B frames. For encoding of this type of frame same process of DCT, Quantization and entropy coding is performed as like happens in JPEG. In DCT, the whole frame is subdivided into 8*8 blocks then DC coefficients are

DPCM coded and after that, the AC coefficients run-length coded. A neat diagram of encoding of I frame is shown in Figure 4.

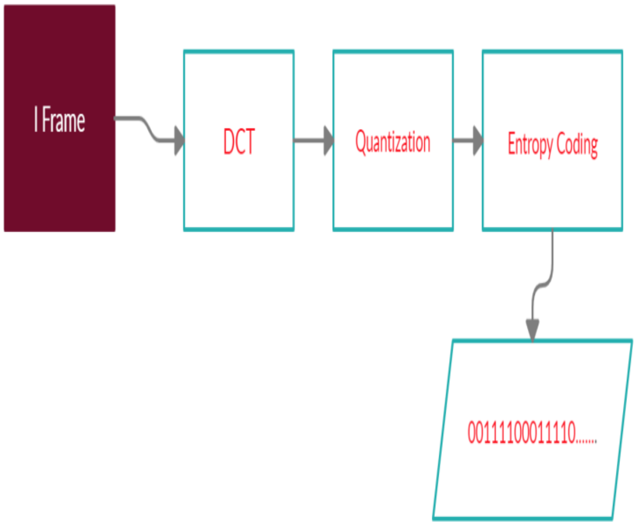

Figure 4. Encoding of I frame

b) Encoding of P Frame: P frame is also known as the forward predicted frame. Firstly the frame is predicted through the corresponding I frame. The search strategy used for finding the motion vector is pentagon search that gives the high performance for motion vector finding probability. Through motion estimation and motion compensation the motion vectors and the errors are basically encoded. To predict the motion of a particular block, a sufficiently larger size window is taken. The corresponding block is searched within the window. The exact match of the block in current, as well as reference frame, is calculated through the different metrics. The position shift of the corresponding block is known as the motion vector. A lot of searched techniques are developed for detecting the motion vector. In our paper pentagon search is used for this purpose that gives better motion vector finding probability than previously proposed search techniques. As fewer blocks movement and inclusion of new block information occurs at this stage. So new information is encoded and the remaining motion vector information is stored in the buffer. The same process of DCT, Quantization and entropy coding is performed for such blocks. Finally at the end of the encoding the visual information is converted into the bitstream. The compression of P frames gives a compression ratio of 20 to 30:1.

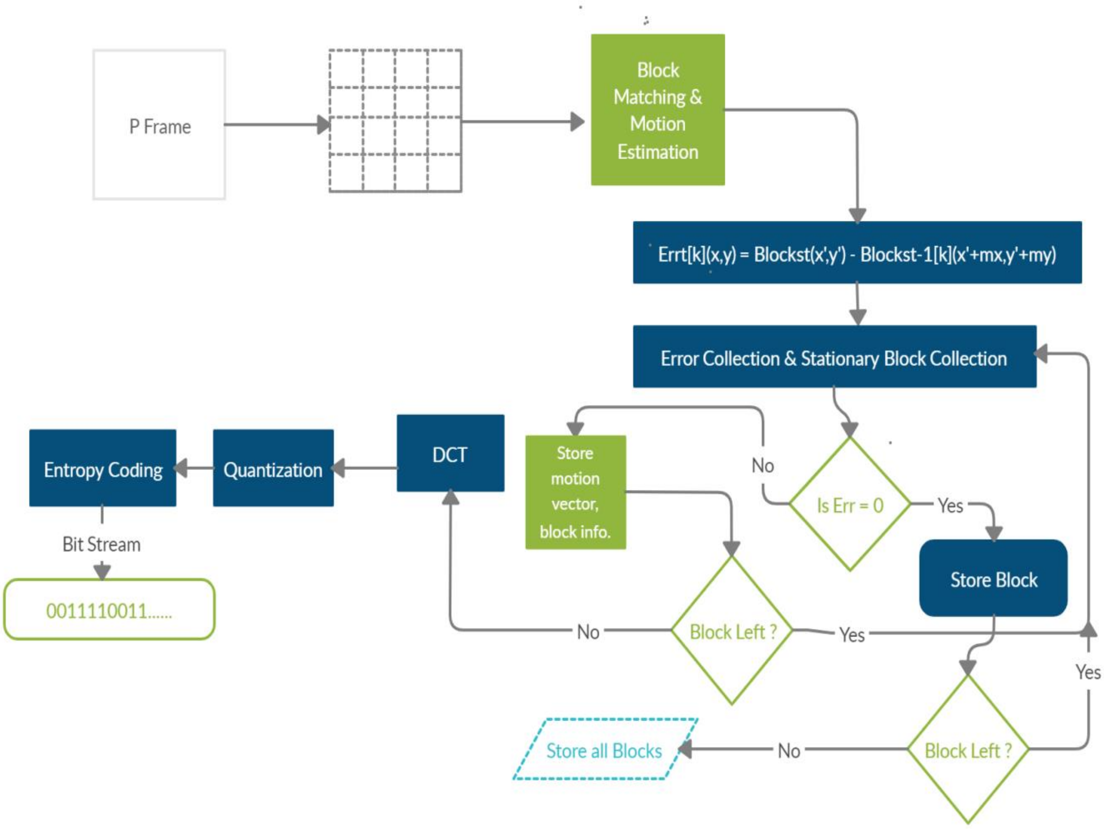

Figure 5. Encoding of P frame

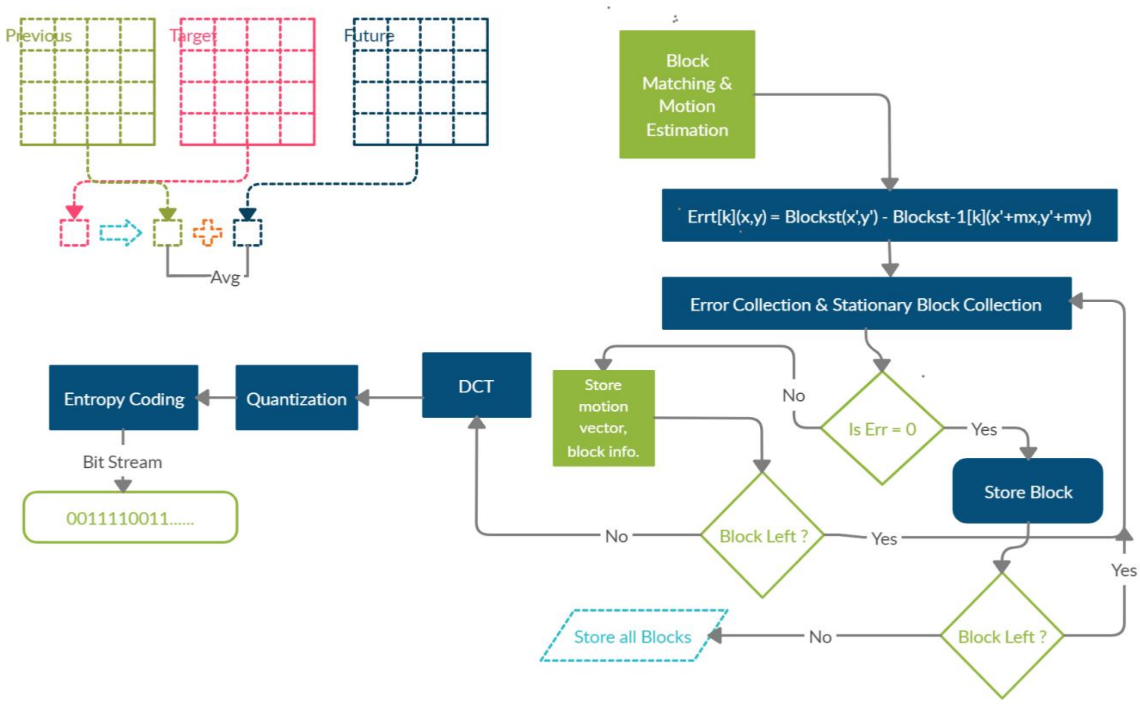

Figure 6. Encoding of B frame

c) **Encoding of B frame:** B frame is also known as the bi directional frame. It can be predicted through the past frame or future frame. All the process of block matching, motion estimation and motion compensation takes place just like as of P frames. At first, suffiently large window size is taken for identifying the motion vector. The search strategy used for finding the motion vector. Firstly the six blocks are chosen for comparison and the errors are calculated. We have already analyzed the pentagon search for finding the motion vector so that it can be used as the most effective search technique for determining the motion vector. There is a slight change in the block matching process in B frame. As in P frames, only past frame is considered as the reference but here past and future both frames are taken into consideration. So, in block matching phase average of these two is taken. Thus compression of B frame takes more space in the buffer. It typically requires very fewer bits than that of I and P frames. The compression ratio of B frames is very high and nearly about 30 to 50:1.

3.2 Decompression strategy of frames: In general the decoding of I frame is just like the decoding of the still jpeg image. For P and B frames there is a slight variation in this process. In the encoding of P and B frames there is a subtraction between input and predicted image, therefore at the decompression end, it is added to get the reconstructed frame. Exact reconstruction of a frame can't be possible because in all cases, at the encoder end, quantization is used that is an irreversible process. So, one has to satisfy at a faithful reconstruction level. Figure 7 shows the diagram of the video decoding. After getting the encoded bit stream, entropy decoding, dequantization, and inverse DCT are performed in order to get the error video. After the error is decoded the predicted video is added for getting the reconstructed frame. Decoding of I, P and B frames are discussed separately. Graphically the process of decoding of I, P, B frames are shown in Figure 8, Figure 9 and Figure 10 respectevily.

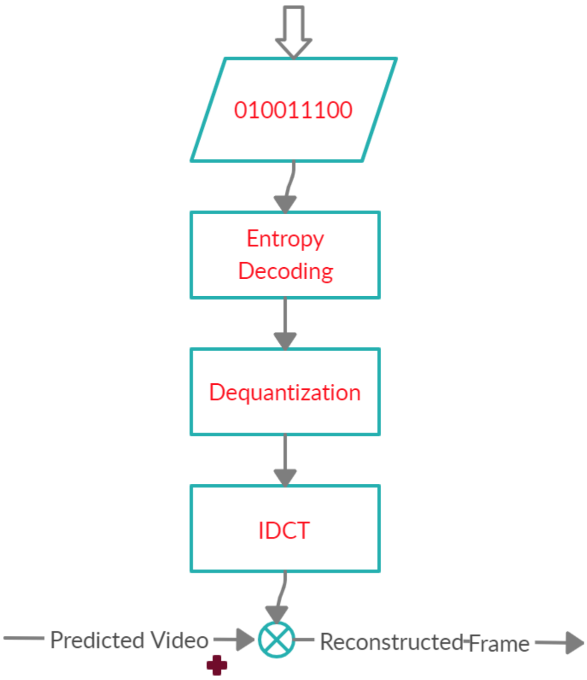

Figure 7. Decompression strategy of Frames

a) **Decoding of I frame:** I frame is decoded in the same manner as the still picture (JPEG) decoding technique. Simply the encoded bit stream gets decoded by entropy decoder, dequantizer and IDCT. The I frame is also known as the Error free frame because it is fully intra coded i.e. no past or future frame is required as the reference frame.

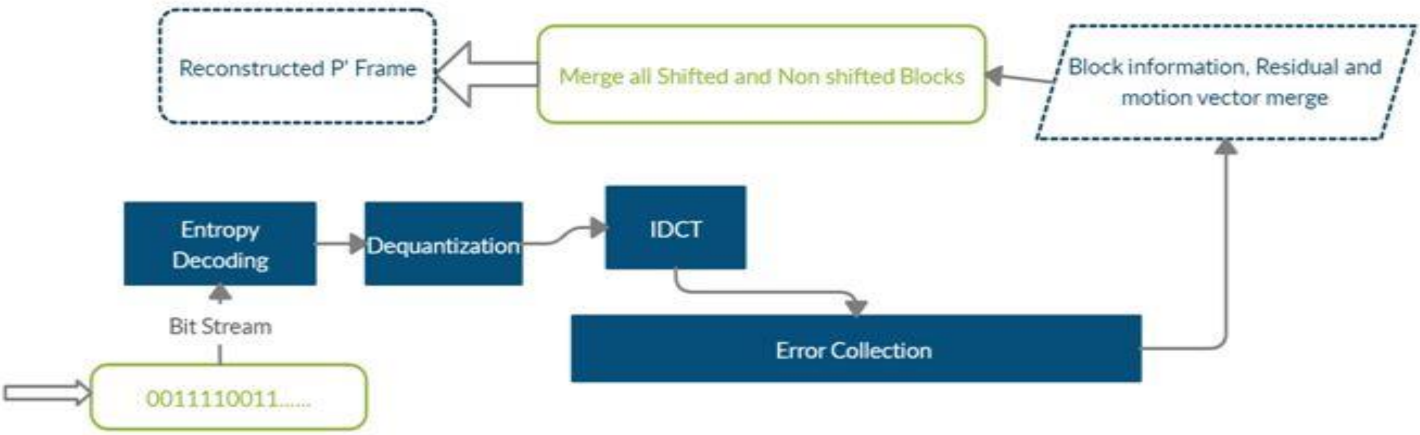

Figure 9. Decompression of P frame

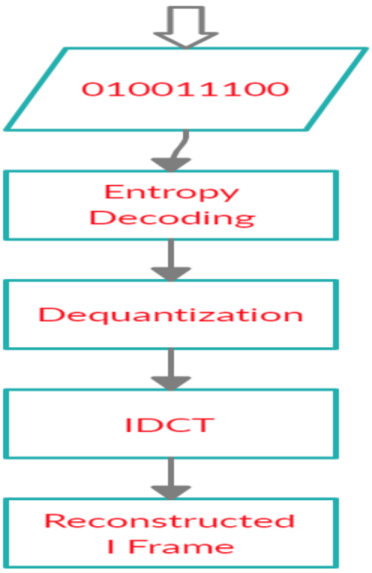

Figure 8. Decompression of I frame

b) Decoding of P frame:

At first the bit stream received at the decoder end, it is firstly entropy decoded followed by dequantization and inverse DCT. After all the above steps whatever received is only the error signal. For a faithful reconstruction, the corresponding motion vector and unchanged block must be appended to the error signal. After merging of the error signal, motion vector, and unchanged block the p frame is reconstructed at the decoder end. It is not actually the P frame but P' frame because at the encoder end there is a step of quantization which is a lossy process, so exact reconstruction is never possible. That's why one has to satisfy only with at an extent of faithful reconstruction. The decoding process of P frame is shown in Figure 9.

c) Decoding of B frame:

Same process of decoding takes place for B
frame as that of P frame. A little bit difference occurs after the error accumulation, as there is only one anchor used as a reference while in the case of B frame two frames are used as a reference frame. The decoding process of B frame is shown in Figure 10.

There are some advantages of coding of B frames that reflects in coding efficiency. B frames has the less but quality information about the motion vector information. Rate of quality improves with accurate prediction and motion estimation. Backward prediction as well as forward prediction involves in the determination of motion vector so for decoding time the correct reference frame plays a major role. The level of accuracy may vary if the combination of forward and backward prediction wrongly used.

There are some disadvantages of handling the B frames. One is directly related to the memory. As it is the prediction of two reference frames, so it requires extra space in memory. In general the memory requirement gets doubled while handling the B frames, so it is not feasible. Another disadvantage appears in case of video conferencing. There is an unwanted delay occurs in this type of real time application.

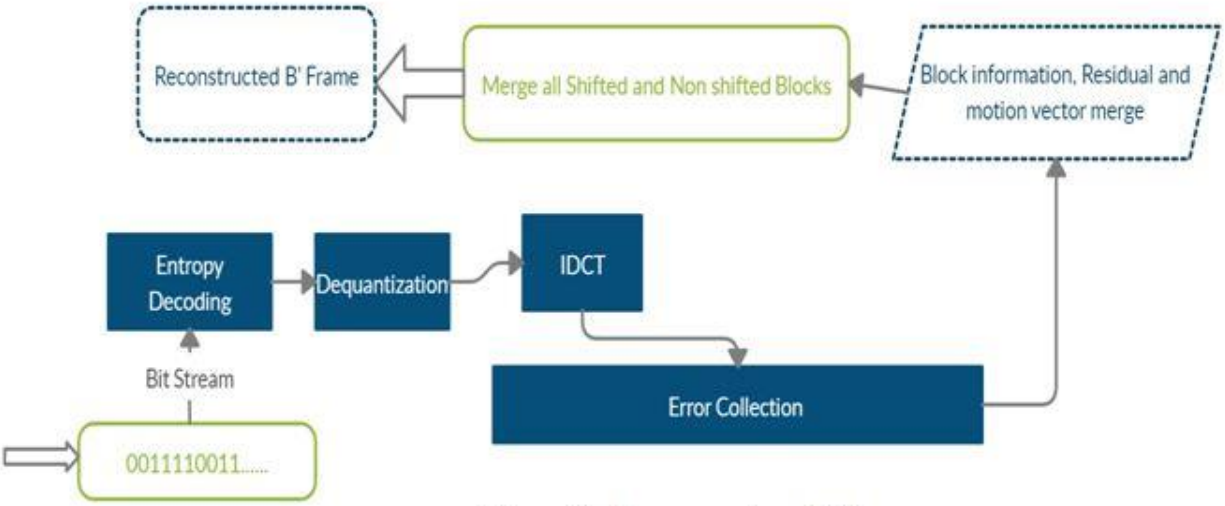

Figure 10. Decompression of B frame

4. Result and Discussion:

As like previous, the performance analysis can be performed through MAD. Our proposed framework efficiently works on P and B frames. We 'he already analyzed that the search pattern of the pentagon method for block matching motion estimation gives a satisfactory performance over different video frames. The search strategy used in this method is pentagon search that searches the motion vector in an efficient manner. Motion vector finding the probability through pentagon search is better than the previously proposed search strategy. Accumulating the error and process through this method is a different idea in both encoding and decoding. It also provides a better simulation result compared to others. The compression performance achieved for P and B frames through this method is larger than any other previously proposed framework. This framework is also applicable for compressing the real time video data.

5. Conclusion

This paper presents a more optimized framework of video compression using pentagon search for block matching motion estimation. As the search, for motion vector takes less time with more accuracy, it can be directly used for the optimized framework of video compression. A slight chane in the mechanism of compression is used in this paper. Whole bunch of blocks (changed and non changed) are processed parallely to reduce the effective time of compression. The same strategy of compression can be used in different domain e.g wavelet domain in future.

References

[1]. P. B. Penafiel and N. M. Namazi, "A new framework for noise-resistant video compression using motion-compensated prediction," *1995 International Conference on Acoustics, Speech, and Signal Processing*, Detroit, MI, USA, vol.4, pp. 2181-2184, 1995.

[2]. K. Minoo and T. Nguyen, "Entropy Coding via Parametric Source Model with Applications in Fast and Efficient Compression of Image and Video Data," *2009 Data Compression Conference*, Snowbird, UT, pp. 461-461, 2009.

[3]. L. Xiaoli, L. Chao, W. Qiang and D. Wenrui, "Design and Implementation of Real-time Video Compression System Control Framework," *2009 WRI International Conference on Communications and Mobile Computing*, Yunnan, pp. 581-585 2009.
[4]. M. Mody, "Video codec framework (VCF): Novel firmware architecture for video hardware," *2014 Twentieth National Conference on Communications (NCC)*, Kanpur, pp. 1-5, 2014.
[5]. S. Moni and S. Sista, "A framework for application specific image compression," *Proceedings DCC '97. Data Compression Conference*, Snowbird, UT, USA, pp. 456-1997.
[6]. Xuguang Yang and K. Ramchandran, "A low-complexity region-based video compression framework using morphology," *Proceedings of 3rd IEEE International Conference on Image Processing*, Lausanne, Switzerland, vol.2, pp. 485-488, 1996.
[7]. L. Herranz and J. M. Martínez, "A Framework for Scalable Summarization of Video," in *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 20, no. 9, pp. 1265-1270, Sept. 2010.
[8]. J. Nightingale, Q. Wang and C. Grecos, "HEVStream: a framework for streaming and evaluation of high efficiency video coding (HEVC) content in loss-prone networks," in *IEEE Transactions on Consumer Electronics*, vol. 58, no. 2, pp. 404-412, May 2012.
[9]. D. Schonberg, S. C. Draper, C. Yeo and K. Ramchandran, "Toward Compression of Encrypted Images and Video Sequences," in *IEEE Transactions on Information Forensics and Security*, vol. 3, no. 4, pp. 749-762, Dec. 2008.
[10]. F. Kamisli, "Recursive Prediction for Joint Spatial and Temporal Prediction in Video Coding," in *IEEE Signal Processing Letters*, vol. 21, no. 6, pp. 732-736, June 2014.
[11]. C. Zhang, K. Ugur, J. Lainema, A. Hallapuro and M. Gabbouj, "Video Coding Using Spatially Varying Transform," in *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 21, no. 2, pp. 127-140, Feb. 2011.
[12]. Neetish Kumar, and Deepa Raj. "A New Pentagon Search Algorithm for Fast Block-Matching Motion Estimation*." International Journal of Innovative Technology and Exploring Engineering (IJITEE)* Volume-8 Issue-10, August 2019