Review of Effective Video Compression Techniques

Yash Salunkhe, Rutuja Jagtap, Harshada Karve, Swaroop Daund
UG Student
Computer Engineering, Zeal College Of Engineering, Pune, India

---

Abstract - In recent few decades digital video compression technologies have become an necessity for the creation, communication and consumption of visual information. Users today have gotten used to taking and posting photos and videos to record daily life, share experiences. According to recent reports, instagram users have been posting an average of 55 million photos and videos every day. How to store, backup and maintain these amount of photos and videos in an efficient way has become an urgent problem. Video compression techniques plays an vital role in storage and transmission of data through the limited bandwidth capability. This review paper introduces advance technology in video compression technique by providing similar quality and perceiving more content of video. This study proposes a compression method that can obtain a highest compression ratio and high quality for video compression by using Discrete cosine transform (DCT), Motion estimation (ME), Embedded zero wavelet (EZW) algorithms.

Keywords - Video compression, Discrete Cosine Transform, Motion estimation Embedded zero wavelet.

---

1. Introduction :

Digital video is a representation of moving visual images in the form of encoded digital data. Digital video comprises a series of digital images displayed in rapid succession. Video compression is the process of encoding a video file in such a way that it consumes less space than the original file and is easier to transmit over the network/internet. It is a type of compression technique that reduces the size of video file formats by eliminating redundant and non-functional data from the original video file.

2. Video compression :

There are two types of videocompression :

A) Lossy Compression :

Lossy compression is the class of data encoding methods that uses inexact approximations and partial data discarding to represent the content. This technique is used to reduce size for storage, handling and transmitting content.

B) Lossless Compression :

Lossless compression reduces a file's size with no less of quality. It is a class of data compression algorithm that allows the original data to be perfectly reconstructed from the compressed data. Lossless compression is used in many applications. For example, it is used in zip file format.

3. Need Of Compression :

Now a day every mobile comes with HD cameras. Even if you take 5 min video, your video file size will be around 200 MB or more. When you want to send this file through facebook or whats app, it will not accept or it will take too much time and bandwidth. It must to reduce video file size.

High bit rates that result from the various types of digital video make their transmission through their intended channels very difficult. Even entertainment video with modest frame rates and dimensions would require bandwidth and storage space far in excess of that available from CD-ROM. Thus delivering consumes quality video on compact disk. Even if high bandwidth technology(example, fiber optic cable) was in place, the per byte cost of transmission would have to be very low before it would be feasible to use it for the staggering amount of data require by HDTV. Finally, even if the storage and transportation problems of digital video were overcome, the processing power needed to manage such volumes of data would make the receiver hardware very expensive. Although significant gain in storage, transmission and processor technology have been achieved in recent years, it is primarily the reduction of amount of data that needs to be stored, transmitted and processed that has made wide spread use of digital video a possibility. This reduction of bandwidth has been made possible by advance in compression technology.

4. DCT [Discrete Cosine Transformation] :

DCT separates images into parts of different frequencies where less important frequencies are discarded through quantization and important frequencies are used to retrieve the image during decompression. Compared to other input dependent transforms, DCT has many advantages:

1. It has been implemented in single integrated circuit.
2. It has the ability to pack most information in fewest coefficients.
3. It minimizes the block like appearance called blocking artifact that results when boundaries between sub-images become visible.

YUV is a color space originally used for Phase Alternating Line (PAL) and composite video colors standards . It is used to convert from RGB to YUV color spaces using the following procedure:

Y = 0.299 R + 0.587 G + 0.114 B (1)

U = 0.492 (B - Y) (2)

V = 0.877 (R - Y) (3)

The YUV model defines a color space in terms of one luminance (Y) and two chrominance (U, V) components. The luminance information can be coded using higher bandwidth than the chrominance information [11], for each n×n data block, a two-dimensional DCT can be written in terms of pixel values f (i, j) for i, j = 0, 1… N-1 and the frequency domain transform coefficients F (u, v) .

4.1 Video Compression Using DCT :

The video compression system consist of video encoder at transmitter side, which encodes the video to be transmitted in form of bits and the video decoder at the receiver side which reconstructs the video in its original form from the bit sequence received. Video consist of number of frames. First frames which are nothing but image sequences. The intra coding method is used for compression of I-frames.

Spatial redundancy :

Elements that are duplicated within a structure such as pixel in a still image and bit pattern in a file. Exploiting spatial redundancy is how compression is performed. Spatial redundancy is redundancy within a single picture or object.

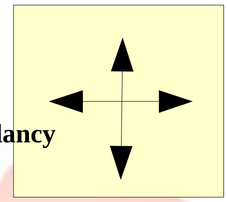

Fig No. 1: Spatial redundancy

- Initially create an I-frames.
- Human eye is insensitive to HF color changes, we convert RGB into luminance and two color different signals that is YUV color model. Y is the luminance (brightness) component while U and V are the chrominance (color) components. We can remove more U,V components than Y.
- DCT value of each pixel is calculated from all other pixel values, so taking 8*8 blocks reduces the processing time.
- The top left pixel in a block is taken as the dc datum for the block. DCT's to the the right of the datum are increasingly higher horizontal spatial frequencies. DCT's below are higher vertical spatial frequencies.
- Using inverse DCT we could reconstruct each pixels value in the 8*8 block.

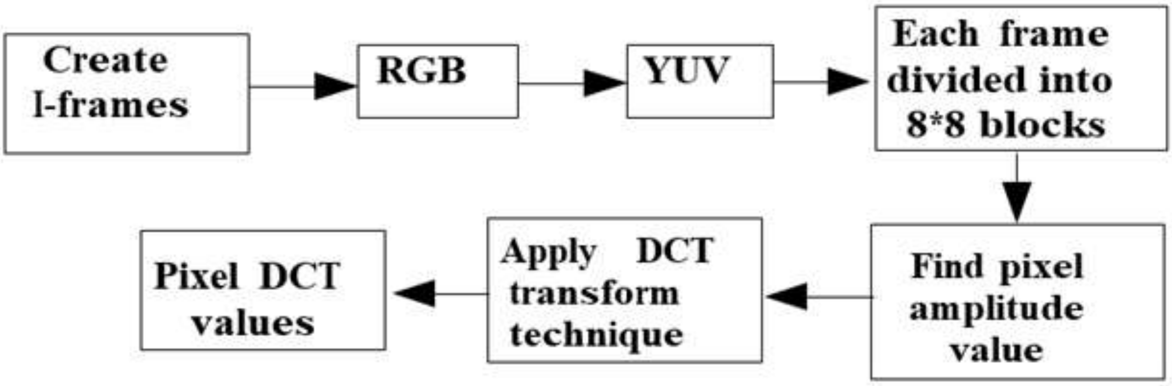

Fig No 2. Spatial redundancy using DCT

4.2 Temporal redundancy :

Pixels in two video frames that have the same values in the same location. Exploiting temporal redundancy is one of the spatial redundancy. The temporal redundancy exists between successive pictures. Adjacent frames are related with each other so can predict one from other and getting residue from them which will give temporal compression in video. According to this, there are three types of frames.

1) I-frame: I-frames are the least compressible but don't require other video frames to decode.
2)) P-frames: P-frames can use data from previous frames to decompress and are more compressible than I-frames.
3) B-frames: B-frames can use both previous and forward frames for data reference to get the highest amount of data compression.

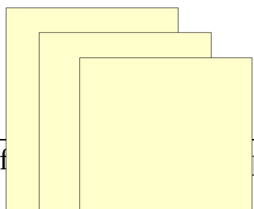

Fig No. 3 : Temporal Redundancy

Inter coded frame is divided into blocks known as macro blocks. After that, instead of directly encoding the raw pixel values for each block, the encoder will try to find a block similar to one it is encoding on a previously encoded frame, referred to as a reference frame. This process is done by a block matching algorithm. If the encoder succeeds on its search, the block could be encoded by a vector known as motion vector, which points to the position of the matching block at the reference frame. The process of motion vector determination is called motion estimation.

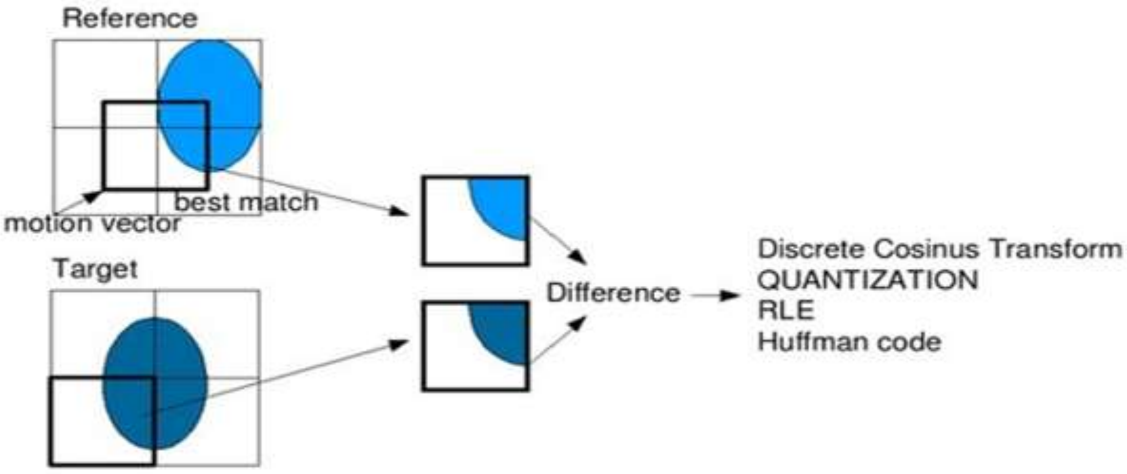

Fig No. 4: Block Matching Algorithm

Following are the steps of Inter frame coding :

1) Frame segmentation: The frame is divided into the number of blocks which are equal in size, non overlapping and rectangular.
2) Frame segmentation: The frame is divided into the number of blocks which are equal in size, non overlapping and rectangular.
3) Block matching: Every blocks in current frame is compare with the every another blocks in past frame. If block size is m and maximum displacements in horizontal and vertical direction are dx and dy respectively, then (2x + m)(2y + m) will be the search area.
4) Block matching: Every blocks in current frame is compare with the every another blocks in past frame. If block size is m and maximum displacements in horizontal and vertical direction are dx and dy respectively, then (2x + m)(2y + m) will be the search area.
5) Motion vector: It describe the location of matching block from past frame with reference to the position of the target block in the current frame.

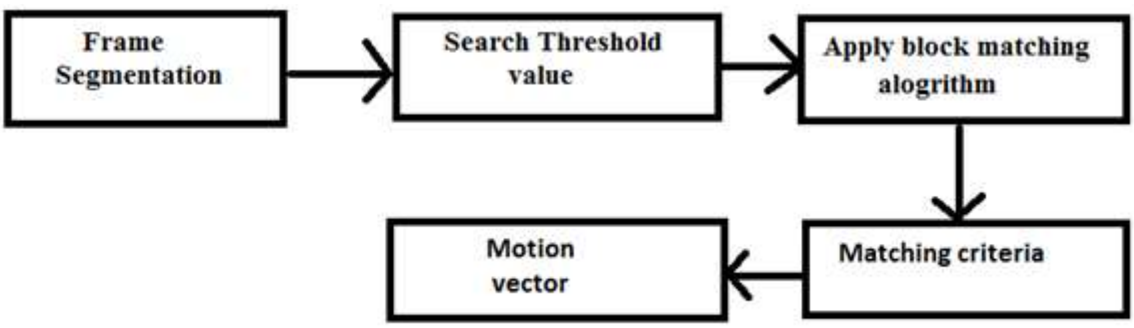

Fig No. 5: Block diagram of inter-frame prediction

4.3 Video Decoder :

1) By using run length decoder bit stream is converted into the original coefficients of DCT.
2) This coefficients are quantized using quantization matrix to original DCT coefficients.
3) To get the approximate component of frames inverse DCT is used.
4) Then by using motion estimation algorithm the frames are motion compensated.
5) YUV is back to converted into the RGB.
6) At last the sequence of the frames are converted into the video which is compressed video.

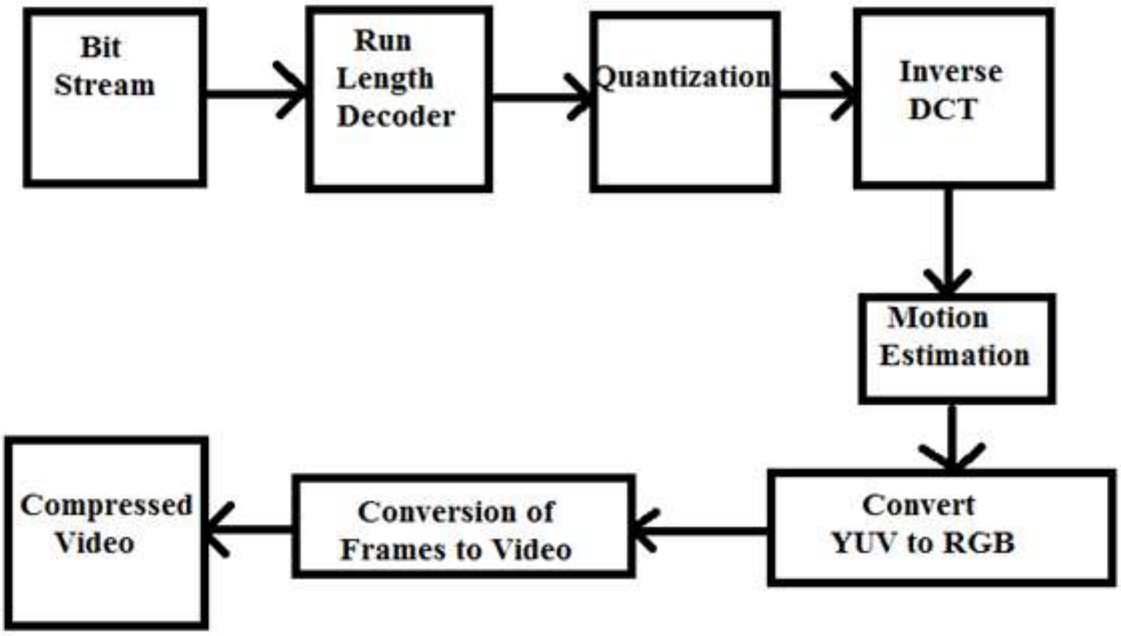

Fig No. 6: Block diagram of Video Decoder

Following formulas are used to measure the degradation after compression :

i. **Peak signal to noise ratio**: PSNR is most easily defined via the mean squared error (MSE). Given a noise-free m×n monochrome image I and its noisy approximation K, MSE is defined as:

$$MSE = \frac{1}{mn}\sum_{i=0}^{m-1}\sum_{j=0}^{n-1}[I(i,j) - K(i,j)]^2$$

The PSNR (in dB) is defined as:

$$\begin{aligned} PSNR &= 10 \cdot \log_{10}\left(\frac{MAX_I^2}{MSE}\right) \\ &= 20 \cdot \log_{10}\left(\frac{MAX_I}{\sqrt{MSE}}\right) \\ &= 20 \cdot \log_{10}(MAX_I) - 10 \cdot \log_{10}(MSE) \end{aligned}$$

ii. **Structural Similarity :**

$$\text{SSIM}(x,y) = \frac{(2\mu_x\mu_y + c_1)(2\sigma_{xy} + c_2)}{(\mu_x^2 + \mu_y^2 + c_1)(\sigma_x^2 + \sigma_y^2 + c_2)}$$

with:

- μ_x the average of x;
- μ_y the average of y;
- σ_x^2 the variance of x;
- σ_y^2 the variance of y;
- σ_{xy} the covariance of x and y;
- $c_1=(k_1L)^2$, $c_2=(k_2L)^2$ two variables to stabilize the division with weak denominator;
- L the dynamic range of the pixel-values (typically this is $2^{\#bits\ per\ pixel}-1$);
- $k_1=0.01$ and $k_2=0.03$ by default.

The SSIM index satisfies the condition of symmetry: $SSIM(x,y) = SSIM(y,x)$

iii. Compressed Ratio :

CR = Size of Original image.
A = (Width * Height * No. of Color Planes * bit – depth) / 8 (bytes)
B =Size of Compressed image = Size in bytes.
Compressed Ratio = A **:** B

5. Motion Estimation in video compression :

Motion estimation based encoders are widely used in video compression techniques, because of its simplicity and coding efficiency. Using model of motion of objects between frames the encoder estimates the motion between two frames that is reference and current frame. This process is called motion estimation. Motion estimation is core and very important component in video processing. Efficiency of video is determined by motion estimation algorithm. Efficiency also depends on coding speed, compression ratio and image quality of decoded video. Motion estimation hard to do as exploration of temporal redundancy between reference and current frame of video sequence allows for significant data compression.

Pixel difference (PD) can be noisy, as noise is there in the sequence can cause problem which can be overcome by smoothing pixel difference before thresholding. Motion estimation is the estimation of displacement of image structures from one frame to another frame in the time sequence of 2D image. Changes between frames is are to space time objects. In motion estimation , motion compensation technique is provide better prediction of current frame. The encoder uses motion model and information to move the content of reference frame.

6. Video compression using EZW :

The embedded zero tree wavelet algorithm (EZW) is a simple, yet remarkably compression algorithm, having the property that the bits in the bit stream are generated in order of importance, yielding a fully embedded code.

The EZW coding algorithm is based on following concepts :

1. A discrete wavelet transform or hierarchical sub band decomposition.
2. Prediction of the absence of significant information across scales by exploiting the self-similarity inherent in images.
3. Entropy-coded successive-approximation quantization.
4. Universal lossless data compression, which is achieved via adaptive arithmetic coding.

7. Future Scope :

Enhancement in compression performance can be implemented by introducing new functionalities which can compress and store high quality of 4K videos. As people come up with new ideas the video compression technology can be improved in following fields :

I. Long term memory.
II. Enhanced motion accuracy.
III. Enhanced intra coding.

8. Conclusion :

We can achieve better quality of images and video by using the approaches like DCT,EZW, RLE algorithms. These approaches will be useful for compressing high memory files and maintaining the quality of the image and video.

References :

[1] Samaher Al-Janabi, Ibrahim Al-Shourbaji, "A Smart and Effective Method for Digital Video Compression"2016 7th International Conference on Sciences of Electronics, Technologies of Information and Telecommunications (SETIT) (2016): 532-538.

[2] Hou, Xingsong, Min Han, Chen Gong, and Xueming Qian. "SAR complex image data compression based on quadtree and zerotree coding in discrete wavelet transform domain: A comparative study." Neurocomputing 148 (2015): 561-568.

[3] Rufai, Awwal Mohammed, GholamrezaAnbarjafari, and Hasan Demirel."Lossy image compression using singular value decomposition and wavelet difference reduction." Digital Signal Processing 24 (2014): 117-123.
[4] Saroya, navpreet; Kaur, Prabhpreet. Analysis of Image Compression Algorithm using DCT and DWT Transforms. International Journal of Advanced Research in Computer Science and Software Engineering,(2014), 4.2: 897-900.
[5] S. H. Ali, "Miner for OACCR: Case of medical data analysis in knowledge discovery," Sciences of Electronics, Technologies of Information and Telecommunications (SETIT), 2012 6th International Conference on, Sousse, 2012, pp. 962-975. doi: 10.1109/SETIT.2012.6482043.
[6] Ram, Idan, Israel Cohen, and Michael Elad. "Facial image compression using patch-ordering-based adaptive wavelet transform." Signal Processing Letters, IEEE, no. 10 (2014): 1270-1274.
[7] Avci Devrya Mustafa Poyraz and Mehmet Kemal Leblebicioglu. " An expert system based on Discrete Wavelet Transform-ANFIS for acquisition and recognition of invariant features from texture images" In 23th Conference on Signal Processing and Communications Applications(SIU),IEEE,(2015):1070-1073.