ISSN: 2278 – 909X
International Journal of Advanced Research in Electronics and Communication Engineering (IJARECE)
Volume 6, Issue 6, June 2017

Video Compression by Key Frame Extraction Based on Entropy Evaluation

Nivedita Deshpande, Kishore Sahoo

***Abstract*— The proposed work projects a technique of video compression using key frame extraction. Since the data compression enhances the utility of transmission channel . Faster transmission speed and low storage has become the need of today's technology. The presented work provides a technique of video data compression which is based on the entropy evaluation and the pixel count of the extracted image. For the Key frame extraction the mean RGB information and entropy is evaluated. The threshold calculation is dependent on these two parameters. The frames are extracted from the video using a reader object. We can summarize the video using the proposed technique and evaluate some parameters like compression ratio, compressed frame count, entropy, average information etc.**

***Index Terms*—Video Compression, Entropy, Key Frame Extraction**

I. Introduction

A very large amount of usage of digital multimedia data through communications, internet, wireless communication and other networks has increased the growth of transmission and processing of data . During signal processing or transmission the data compression is required which involves the source coding or bit rate reduction. This involves encoding the information using fewer bits as compare to the original representation. The process of reducing the size of a data file is known as data compression. Compression is useful since it reduces resources(memory) required to store and transmit data. Data compression involves the compression of text signals, image signals, speech signals and video signals.Video signals differ from image signals in several important characteristics. The most important difference is that video signals have a camera frame rate of anywhere from 15 to 60 frames/s, which provides the illusion of smooth motion in the displayed signal. Another difference between images and video is the ability to exploit temporal redundancy as well as spatial redundancy in designing compression methods for video. Most of the video compression algorithms use the coding techniques to reduce the redundancy in the video data. These algorithms combine the spatial image compression and the temporal motion compensation[Ref. no. 1 &2]. Video data may be represented as a series of still image frames. The sequence of frames contains spatial and temporal redundancy that video compression algorithms attempt to eliminate or code in a smaller size. Similarities can be encoded by only storing differences between frames, or by using perceptual features of human vision. For example, small differences in color are more difficult to perceive than are changes in brightness. Compression algorithms can average a color across these similar areas to reduce space. There are many methods used for video compression. A video compression system consists of the following:

An encoder , Compressed bit-streams and a decoder. The most famous and apply technique is Moving Picture Experts Group (MPEG), which is an ISO/ITU standard for compressing digital video. MPEG performs lossy compression for each frame similar to JPEG, which means pixels from the original images are permanently removed. The most famous methods are concentrated of Discrete Cosine Transform (DCT) and Discrete Wavelet Transform (DWT)[1,2]. The grey scale image gives 256 levels of possible intensity for each pixel, so these images refer to 8 bits per pixel (bpp). The typical RGB color images, with 8 bits for Red, 8 bits for Green, and 8 bits for Blue, then the intensity I is defined by (I=R+G+B). The human eye is most sensitive to variations in intensity, so the most difficult part of compressing a color image lies in the compressing of the intensity Compression is a reversible conversion (encoding) of data that contains fewer bits. This allows a more efficient storage and transmission of the data.

II. Video Compression

The MPEG compression algorithm encodes the data in following steps-

First a reduction of the resolution is done, which is followed by a motion compensation in order to reduce temporal redundancy.

The next steps are the Discrete Cosine Transformation (DCT) and a quantization as it is used for the JPEG compression; this reduces the spatial redundancy (referring to human visual perception).

The final step is an entropy coding using the Run Length Encoding and the Huffman coding algorithm. An MPEG video can be understood as a sequence of frames. Because two successive frames of a video sequence often have small differences (except in scene changes), the MPEG-standard offers a way of reducing this temporal redundancy. Motion compensation requires a large amount of computation. There are several ways we can reduce the total number of computations. One way is to increase the size of the block. Increasing the size of the block means more computations per comparison. However, it also means that we will have fewer blocks per frame, so the number of times we have to perform the motion compensation will decrease. However, different objects in a frame may be moving in different directions[3,4,5].

It uses three types of frames: I-frames (intra), P-frames (predicted) and B-frames (bidirectional).

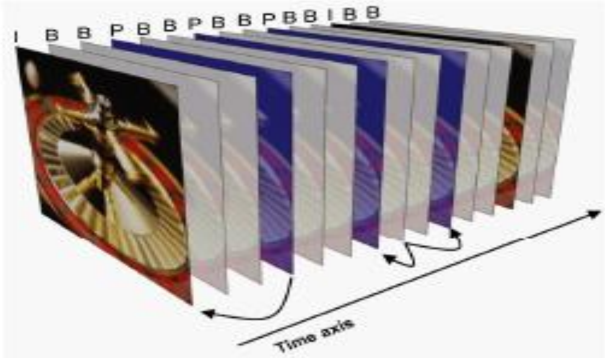

Fig 2.1: An MPEG frame sequence with two possible references: a P-frame referring to a I-frame and a B-frame referring to two P-frames.

The I-frames are "**key-frames**", which have no reference to other frames and their compression is not that high. The P-frames can be predicted from an earlier I-frame or P-frame. P-frames cannot be reconstructed without their referencing frame, but they need less space than the I-frames, because only the differences are stored. The B-frames are a two directional version of the P-frame, referring to both directions (one forward frame and one backward frame). B-frames cannot be referenced by other P- or B-frames, because they are interpolated from forward and backward frames. P-frames and B-frames are called inter coded frames, whereas I-frames are known as intra coded frames. Proposed method uses the key frame extraction[6,7,8].
The references between the different types of frames are realised by a process called motion estimation or motion compensation. The correlation between two frames in terms of motion is represented by a motion vector. The resulting frame correlation, and therefore the pixel arithmetic difference, strongly depends on how good the motion estimation algorithm is implemented.

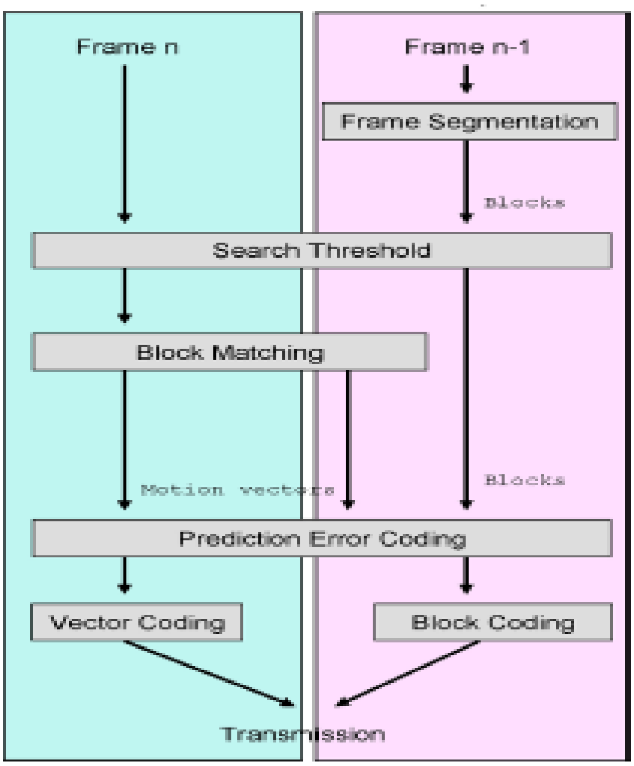

Fig 2.2: Information flow of motion estimation

Prediction Error Coding: Video motions are often more complex, and a simple "shifting in 2D" is not a perfectly suitable description of the motion in the actual scene, causing so called prediction errors . The MPEG stream contains a matrix for compensating this error. After prediction the, the predicted and the original frame are compared, and their differences are coded. Obviously less data is needed to store only the differences (Middle image shown in 2nd row showing difference in yellow and black regions in Fig. 2.3).

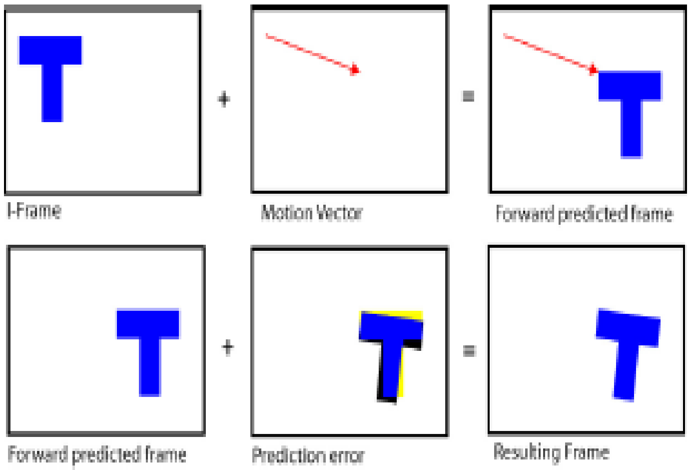

Fig 2.3: Image showing prediction error

Vector Coding: After determining the motion vectors and evaluating the correction, these can be compressed. Large parts of MPEG videos consist of B- and P-frames as seen before, and most of them have mainly stored motion vectors. Therefore an efficient compression of motion vector data, which has usually high correlation, is desired[9-12]. Discrete Cosine Transform (DCT) used in Block coding.DCT allows, similar to the Fast Fourier Transform (FFT), a representation of image data in terms of frequency components[13-14]. So the frame-blocks (8x8 or 16x16 pixels) can be represented as frequency components. The transformation into the frequency domain is described by the following formula:

F(u,v)=

$$\frac{1}{4} C(u)C(v) \sum_{x=0}^{N-1}\sum_{y=0}^{N-1} f(x,y) \cos\frac{(2x+1)u\pi}{2N} . \cos\frac{(2y+1)v\pi}{2N} \quad \text{....(2.1)}$$

$$C(u)C(v) = \frac{1}{\sqrt{2}} \; for\; u,v = 0$$

$$C(u)C(v) = 1, else$$

N = block size

The inverse DCT is defined as:

$$F(u,v)= \frac{1}{4}\sum_{u=0}^{N-1}\sum_{v=0}^{N-1} C(u)C(v)F(u,v)\cos\frac{(2y+1)u\pi}{16} . \cos\frac{(2x+1)v\pi}{16} \quad ...(2.2)$$

III. Calculating Image Entropy

Entropy is a statistical quantity which is used to describe the randomness' of an image, i.e. the amount of information which must be coded for by a compression algorithm. Low entropy images, such as those containing a lot of black sky, have very little contrast and large runs of pixels with the same or similar DN values. An image that is perfectly flat will have an entropy of zero. Consequently, they can be compressed to a relatively small size. On the other hand, high entropy images such as an image of heavily cratered areas on the moon have a great deal of contrast from one pixel to the next and consequently cannot be compressed as much as low entropy images. Image entropy is calculated as-

$$\text{Entropy} = - \sum_{i} P_i \log_2 P_i \quad ....(3.3)$$

In the above expression, P_i is the probability that the difference between 2 adjacent pixels is equal to i, and Log_2 is the base 2 logarithm.

IV. Methodology

The algorithm for the proposed method is applied on various sample videos and some parameters have been evaluated. The flow chart is shown in Figure4.1. The video signal is read as a sequence of frames. These frames are read as an image. Then the entropy and average information of each frame is evaluated . If the entropy and average information of frame is more than the average entropy and mean information then the frame is selected as the key frame. Otherwise next frame will be selected . The compressed video contain the key frames.

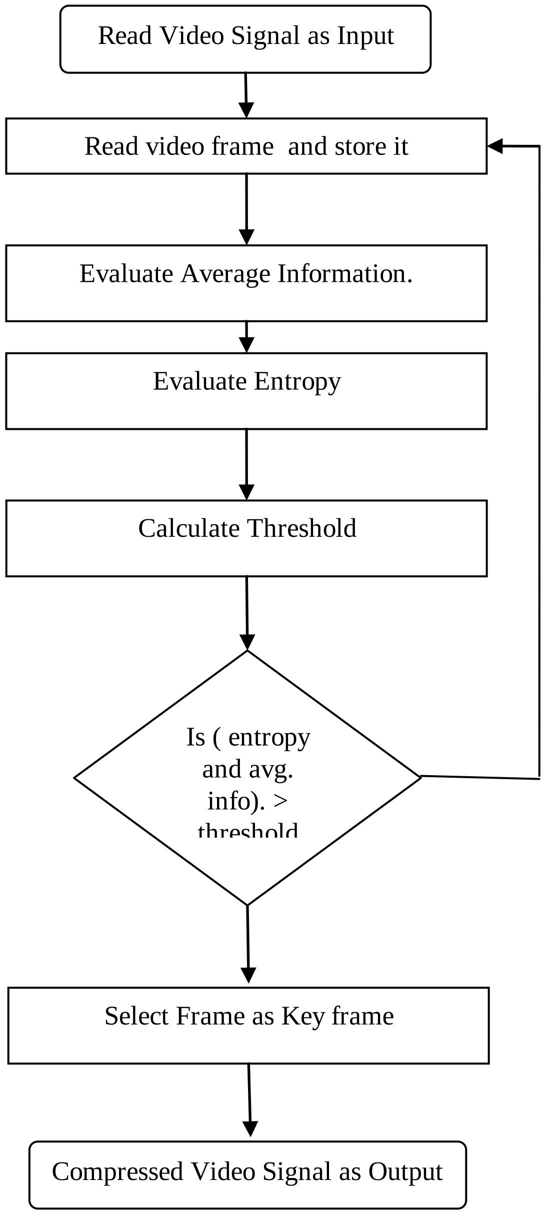

Figure 4.1: Flow chart

V. RESULT & DISCUSSION

A few of the key frames extracted from the original video have been shown below-


(Avg. Info: 196.07) (Avg. Info: 200.47)


(Avg. Info: 200.57) (Avg. Info: 201.23)


(Avg. Info:202.34) (Avg. Info: 203.32)

The video summarization details are tabulated below-

Average Information	194.63
Entropy	17.11
Key Frames	79
Compression Ratio	49.06%
Time Compressed	83.33%

VI. CONCLUSION

Video compression using key frame extraction based on entropy and average information can be done and compressed video can be obtained without much loss of information. The algorithm provides evaluation of various parameters like compression ratio, entropy, average information.

References

[1] Khalid Sayood, “Introduction to Data Compression”, Morgan Kaufmann Publishers is an imprint of Elsevier(2006),Edition III.

[2] Huffman, D. A. (1951),” A method for the construction of minimum redundancy codes” ,Proceedings of the Institute of Radio Engineers ,(1951),40, pp. 1098-1101.

[3] Azra Nasreen et al.”Key Frame Extraction from Videos - A Survey “,International Journal of Computer Science & Communication Networks,(2011),Vol 3(3),194-198

[4] Guozhu Liu et al. ,” Key Frame Extraction from MPEG Video Stream”, Proceedings of the Second Symposium International Computer Science and Computational Technology(ISCSCT), Huangshan, P. R. China, 26-28,Dec. 2009, pp. 007-011.

[5] Sheena C.V. et al.,” Key-frame extraction by analysis of histograms of video frames using statistical methods”, 4’th International Conference on Eco-friendly Computing and Communication Systems(© 2014 The Authors. Published by **Elsevier** B.V. Peer-review under responsibility of organizing committee of the International Conference on Eco-friendly Computing and Communication Systems (ICECCS -2015).

[6] Qiang Zhang et al. ,”An Efficient Method of Key-Frame Extraction Based on a Cluster Algorithm”, J Hum Kinet. (2013 Dec 18), 39: pp5–13, (http://creativecommons.org/licenses/by/3.0/).

[7] S. Ariffa Begum et al.,” Performance Analysis of Various Key Frame Extraction Methods for Surveillance Applications, International Journal of Emerging Technology and Advanced Engineering “,Website: www.ijetae.com (ISSN 2250-2459, ISO 9001:2008 Certified Journal, (September 2014) ,Volume 4, Issue 9, 157 .

[8] Shaoshuai Lei, “A Novel Key-Frame Extraction Approach for Both Video Summary and Video Index’, The Scientific World Journal, (2014),Volume 2014 , Article ID 695168, 9 pages, http://dx.doi.org/10.1155/2014/695168.

[9] Mei Huang et al.,”An Integrated Scheme for Video Key Frame Extraction”, 2nd International Symposium on Computer, Communication, Control and Automation (3CA 2013).

[10] Ganesh I Rathod et al.,”An Algorithm for Shot Boundary Detection and Key Frame Extraction Using Histogram Difference”,International Journal of Emerging Technology and Advanced Engineering Website: www.ijetae.com August 2013, (ISSN 2250-2459, ISO 9001:2008 Certified Journal, Volume 3, Issue 8)

[11] MuzhirShaban Al-Ani et al.,” Video Compression Algorithm Based on Frame Difference Approaches”, International Journal on Soft Computing (IJSC) ,(2011),Vol.2, No.4.

[12] Rajeshwar Dass et al., “Video Compression Technique”, International Journal of Scientific & Technology Research Volume 1, (November 2012), Issue 10, ISSN 2277-8616.

[13] Jaiswal, R.C.. Audio-Video Engineering. Pune, Maharashtra: Nirali Prakashan. p. 3.41. (2009), ISBN 9788190639675.

[14]http://www.videoproductionslondon.com/blog/ scene-change-detection-during-encoding-key-frame-extraction