International Journal of Emerging Trends & Technology in Computer Science (IJETTCS)
Web Site: www.ijettcs.org Email: editor@ijettcs.org
Volume 6, Issue 3, May- June 2017 ISSN 2278-6856

A Study of Video Summarization Using Various Compression Techniques

Nivedita Deshpande[1], Kishore Sahoo[2]

[1].M.Tech. Scholar, Dept. of Electronics & Telecommunication Engg. CCEM, Raipur, (C.G.) India
[2].Head & Prof , Dept. of Electronics & Telecommunication Engg. CCEM, Raipur, (C.G.) India

Abstract: ***This paper presents a study of various video compression techniques. 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 .The compression of video signals is required since it reduces resources required to store and transmit data. 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 .There are two important video and image compression standards JPEG and MPEG. JPEG is associated with still digital pictures while MPEG is associated with the digital video sequence.***

Keywords: Video Compression, MPEG, Frame

1. Introduction

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]. This can be done by implementation of source coding technique. When we speak of a compression technique or compression algorithm as shown in Figure1.1, we refer to two algorithms. There is the compression algorithm that takes an input X and generates a representation $\hat{X}$, which requires fewer bits, and there is a reconstruction algorithm that operates on the compressed representation $\hat{X}$ to generate the reconstruction y.

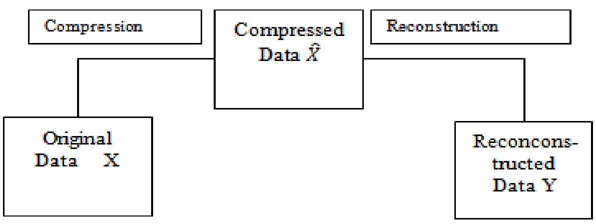

Fig1. 1: Compression Algorithm

The data compression schemes can be divided into two broad classes: *lossless* compression schemes, in which Y is identical to X, and *lossy* compression schemes, which generally provide much higher compression than lossless compression but allow Y to be different from *X*. That is no information is lost in lossless compression. The lossless compression reduces the bits by identifying and eliminating the redundancy[12,3]. Lossy compression reduces bits by removing unnecessary or less important information. The most of the video compression algorithms use lossy compression since the lossy compression has a trade off between the video quality and compression factor. The uncompressed video data requires a very high data rate. For example , a compression technique for video sometimes require costly hardware for the video decompression. The design of data compression schemes involves the trade-offs among various factors, including degree of compression, the amount of distortion introduced and the computational resources required to compress and decompress the data.

A lossless video compression codecs perform at a compression factor of 5-12, while a typical MPEG-4 lossy compression video has a compression factor between 20 and 200. Highly compressed video may present visible or distracting artifacts[1,2,3]. Lossless data compression algorithms usually exploit statistical redundancy to represent data without losing any information, so that the process is reversible. Lossless compression is possible because most real-world data exhibits statistical redundancy. Lossy compression reduces bits by removing unnecessary or less important information. The process of reducing the size of a data file is referred to as data compression. In the context of data transmission, it is called source coding (encoding done at the source of the data before it is stored or transmitted).

2. MPEG Standards

MPEG stands for Moving Picture Coding Exports Group . At the same time it describes a whole family of international standards for the compression of audio-visual digital data. The most known are MPEG-1, MPEG-2 and MPEG-4, which are also formally known as ISO/IEC-11172, ISO/IEC-13818 and ISO/IEC- 14496. More details about the MPEG standards can be found in[Ref. no1,2,4,5]. The most important aspects are summarized as follows:

2.1.MPEG-1

The MPEG-1 Standard was published 1992 and its aim was it to provide VHS quality with a bandwidth of 1,5

Mb/s, which allowed to play a video in real time from a 1x CD-ROM. The frame rate in MPEG-1 is locked at 25 (PAL) fps and 30 (NTSC) fps respectively. The MPEG standard at least initially had applications that require digital storage and retrieval as a major focus.

2.2.MPEG-2

In 1994 MPEG-2 was released, which allowed a higher quality with a slightly higher bandwidth. MPEG-2 is compatible to MPEG-1. Later it was also used for *High Definition Television (HDTV)* and *DVD*, which made the MPEG-3 standard disappear completely. The frame rate is locked at 25 (PAL) fps and 30 (NTSC) fps respectively, just as in MPEG-1. MPEG-2 is more scalable than MPEG-1 and is able to play the same video in different resolutions and frame rates.

2.3.MPEG-4

Digital video consists of a stream of images captured at regular time intervals. The images are represented as digitized samples containing visual (color and intensity) information at each spatial and temporal location. Visual information at each sample point may be represented by the values of the three basic color components RGB color space. A video signal can be sampled in either frames (progressive) or fields (interlaced). In progressive video, a complete frame is MPEG-4 was released 1998 and it provided lower bit rates (10Kb/s to 1Mb/s) with a good quality. It was a major development from MPEG-2 and was designed for the use in interactive environments, such as multimedia applications and video communication.

The latest video compression standard, H.264 (also known as MPEG-4 Part 10/AVC for Advanced Video Coding), is expected to become the video standard of choice in the coming years.

3.1.Video Compression Techniques

There are various different techniques available for video compression. The techniques are based on key frame extraction. The video file is converted into frames and then summarization technique is applied[6,7,8-10].

3.1.1Key Frame Extraction

The video file is converted in to a sequence of frames and reduction of resolution is done. 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.

The frames are categorized as I- frame, P-frame, B-frame. 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.

3.1.2.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. The usage of the particular frame type defines the quality and the compression ratio of the compressed video. I-frames increase the quality (and size), whereas the usage of B-frames compresses better but also produces poorer quality. The distance between two I-frames can be seen as a measure for the quality of an MPEG-video. In practise following sequence showed to give good results for quality and compression level: **I***BB*P*BB*P*BB*P*BB***I**BBP.

3.1.3.Frame Segmentation

The Actual frame is divided into non-overlapping blocks (macro blocks) usually 8x8 or 16x16 pixels. Motion compensation requires a large amount of computation. several ways we can reduce the total number of computations; as one way is to increase the size of the block.

3.1.4.Block Matching

In general block matching tries, to "stitch together" an actual predicted frame by using snippets (blocks) from previous frames. The process of block matching is the most time consuming one during encoding. In order to find a matching block, each block of the current frame is compared with a past frame within a search area. Only the luminance information is used to compare the blocks, but obviously the colour information will be included in the encoding. The search area is a critical factor for the quality of the matching. It uses histogram difference technique or mean square error technique[11,12,13].

Image histogram is a type of histogram that acts as a graphical representation of the tonal distribution in a digital image[1] .

Mean Square method is used to find the dissimilarity D(s,t) (sometimes referred to as error, distortion, or distance) between two images n and n-1 is defined as follows-

$$D(s,t) = \sum_{Vy=1}^{p} \sum_{Vx,=1}^{q} M[\Psi_n(x,y), \Psi_{n-1}(x+V_x, y+V_y)]$$

where **M(u,v), Mean Square Error** is a metric that measure the dissimilarity between the two arguments u and v.

(MSE): 2 $M(u, v) = (u - v)^2$

3.1.5.Search Threshold

In order to minimize the number of expensive motion estimation calculations, they are only calculated if the difference between two blocks at the same position is higher than a threshold, otherwise the whole block is transmitted[13,14,15].

3.1.6.Discrete Cosine Transform (DCT)

DCT is used in Block coding.DCT allows, similar to the Fast Fourier Transform (FFT), a representation of image data in terms of frequency components[15, 16]. 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:

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

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

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

N = block size

The inverse DCT is defined as:

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

3.1.7.Quantization and Entropy Coding

During quantization, which is the primary source of data loss, the DCT terms are divided by a quantization matrix, which takes into account human visual perception. The transform coefficients are then quantized. The function of quantization matrix is to quantize the high frequencies with coarse quantization steps because the human visual perception is less sensitive to the high frequencies. The bits are saved for coding high frequencies are used for low frequencies[17,18,19].

The entropy coding takes two steps: Run Length Encoding (RLE) [Ref. no.1,18] and Huffman coding [Ref. no. 1]. These are well known lossless compression methods, which can compress data, depending on its redundancy.

5. Conclusion

This paper summarizes different methods of video compression . The extraction of key frame is the basic task for the video compression, which itself involves various techniques. The role of key- frame extraction is to reduce redundant frames that can lead dimensionality reduction of feature vector for classification.

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] PEREIRA, F. The MPEG4 Standard: Evolution or Revolution www.img.lx.it.pt/~fp/artigos/VLVB96.doc (3. Feb. 2006)

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

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

[15] Faxin Yu; Hao Luo; Zheming Lu.”Three-Dimensional Model Analysis and Processing”, Berlin: (2010)Springer. p. 47. ISBN 9783642126512.

[16]Ahmed, N.; Natarajan, T.; Rao, K.R.. “Discrete Cosine Transform”. IEEE Transactions on Computers. (January 1974), **C–23** (1): 90–93. doi:10.1109/T-C.1974.223784.

[17]Michael Freeman , “The Digital SLR Handbook”. (2005),Ilex. ISBN 1-904705-36-7.

[18] Capon, J. .” A probabilistie model for run-length coding of pictures”, IRE Trans. On Information Theory, (1959), IT-5, (4), pp. 157-163.

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