ISSN: 2321-7782 (Online)
Volume 2, Issue 8, August 2014
International Journal of Advance Research in Computer Science and Management Studies
Research Article / Survey Paper / Case Study
Available online at: www.ijarcsms.com

Key Frame Extraction using wavelet transforms – A Video Summarization Technique

Chethan Sharma
Department of CSE
CUFE
Bangalore – India

Abstract: In current era, most of the information is captured using multimedia techniques. Most used methods for information capturing is through images and videos. In processing a video, large information needs to be processed and a number of frames could contain similar information which could cause unnecessary delay in gathering the required information. Video summarization can speed up video processing. There are different techniques for video summarization. In this paper key frames are used for summarization. Key frames are extracted using discrete wavelet transforms as key frame extraction by DWT is found to be more efficient than DCT.

Keywords: Video Processing, wavelet, DWT, Key Frames, Video Summarization.

I. INTRODUCTION

Digital images and videos are everywhere these days – in thousands of scientific (e.g., astronomical, bio-medical), consumer, industrial, and artistic applications. Moreover they come in a wide range of the electromagnetic spectrum - from visible light and infrared to gamma rays and beyond. The ability to process image and video signals is therefore an incredibly important skill to master for engineering/science students, software developers, and practicing scientists. Digital image and video processing continues to enable the multimedia technology revolution we are experiencing today. Some important examples of image and video processing include the removal of degradations images suffer during acquisition (e.g., removing blur from a picture of a fast moving car), and the compression and transmission of images and videos, for economical storage and efficient transmission [1].

Enormous popularity of the Internet video repository sites like YouTube, Yahoo Video, lecture videos, and social networking sites like face book, Google+ etc. have caused increasing amount of the video content available over the Internet. In such a scenario, it is necessary to have automatic mechanisms of generating concise representation of the video content as a sequence of still or moving pictures i.e. video summary. The major task in video summarization is to segment the original video into shots and extract those video frames from the original video that would be the most informative and concise representation of the whole video.

Such frames are referred as key frames [2]. Fig. 1 represents the anatomy of a video. Key frames can be extracted locally or globally using various visual or audio features [2, 3, 4].

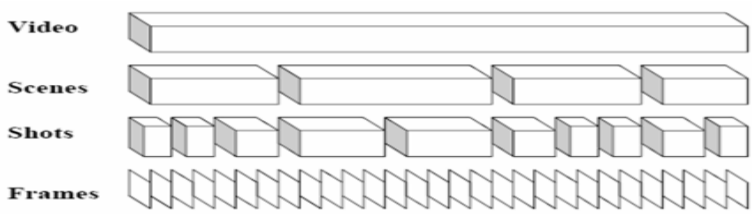

Fig. 1 Anatomy of a video

Video summarization is a mechanism for generating a short summary of a video, which can either be a sequence of stationary images (key frames) or moving images (video skims) [5]. Video can be summarized by two different ways which are as follows.

1) Key Frame Based Video Summarization

These are also called representative frames, R-frames, still-image abstracts or static storyboard, and a set consists of a collection of salient images extracted from the underlying video source [5]. Following are some of the challenges that should be taken care while implementing Key frame based algorithm

i. Redundancy: frames with minor difference are selected as key frame.

ii. When there are various changes in content it is difficult to make clustering.

2) Video Skim Based Video Summarization

This is also called a moving-image abstract, moving story board, or summary sequence [5]. The original video is segmented into various parts which is a video clip with shorter duration. Each segment is joined by either a cut or a gradual effect. The trailer of movie is the best example for video skimming [6].

II. Proposed method-Key Frame Extraction

Video contains huge amount of information at different levels in terms of scenes, shots and frames. To discover knowledge from videos the issue that needs to be addressed is the elimination of redundant information. The Objective is to remove the redundant data which will significantly reduce the amount of information that needs to be processed. So, key frame extraction is the fundamental step in any of the video retrieval applications. It is necessary to discard the frames with repetitive or redundant information during the extraction.

In recent years, many algorithms of key frame extraction focused on original video stream have been proposed. Key frame is the frame which can represent the salient content of the shot. The key frames extracted must summarize the characteristics of the video, all the key frames on the time sequence gives visual summary of the video to the user. There are great redundancies among the frames in the same shot, so only those frames that best reflect the shot contents are selected as key frames to represent the shot. The extracted key frames should contain as much salient content of the shot as possible and avoid as much redundancy as possible [7].

Two most used techniques for compressing is DCT and DWT. In this paper, DWT is considered for compressing the video because of following reasons [8]

1) DWT provides higher compression ratios
2) Avoids blocking artifacts.
3) Allows good localization both in spatial & frequency domain.
4) Transformation of the whole image introduces inherent scaling.
5) Better identification of which data is relevant to human perception higher compression ratio and
6) DCT takes more time than DWT.

Wavelet Transform has become an important method for image compression. Wavelet based coding provides substantial improvement in picture quality at high compression ratios mainly due to better energy compaction property of wavelet transforms [9]. Wavelets are functions which allow data analysis of signals or images, according to scales or resolutions. The DWT represents an image as a sum of wavelet functions, known as *wavelets*, with different location and scale. It represents the data into a set of high pass (detail) and low pass (approximate) coefficients. The input data is passed through set of low pass and high pass filters. The output of high pass and low pass filters are down sampled by 2 .The output from low pass filter is an approximate coefficient and the output from the high pass filter is a detail coefficient [10]. This procedure is one dimensional (1-D) DWT but in this research work we are using two dimensional (2-D) DWT. The outputs are then down sampled by 2 in each direction as in case of 1-D DWT [9]. Output is obtained in set of four coefficients LL, HL, LH 2-D DWT, the input data is passed through set of both low pass and high pass filter and HH. The first alphabet represents the transform in row where as the second alphabet represents transform in column. The alphabet L means low pass signal and H means high pass signal. LH signal is a low pass signal in row and a high pass in column. Hence, LH signal contain horizontal elements. Similarly, HL and HH contains vertical and diagonal elements, respectively [8,11] . Example for 2D-DWT is shown in Fig 2.

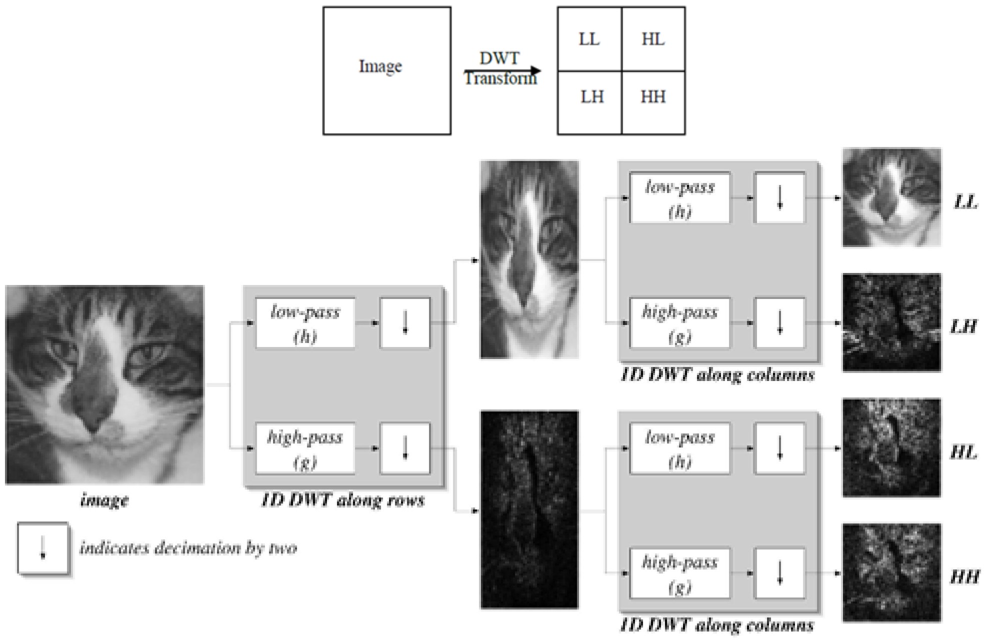

Fig. 2 2D-DWT

Fig.3 shows the Design of Key Frame Extraction. For each frame DWT is performed to get the DWT sub bands. Each of the LL, LH, HL and HH sub bands are obtained performing operations on the image. The sub bands thus obtained are used to update the corresponding mean and standard deviations. The threshold is also updated and the deviations are compared with threshold to determine if the current frame is a key frame or not. If deviation of one of the sub bands of the image is greater than the threshold for that sub band, then it is a key frame [14].

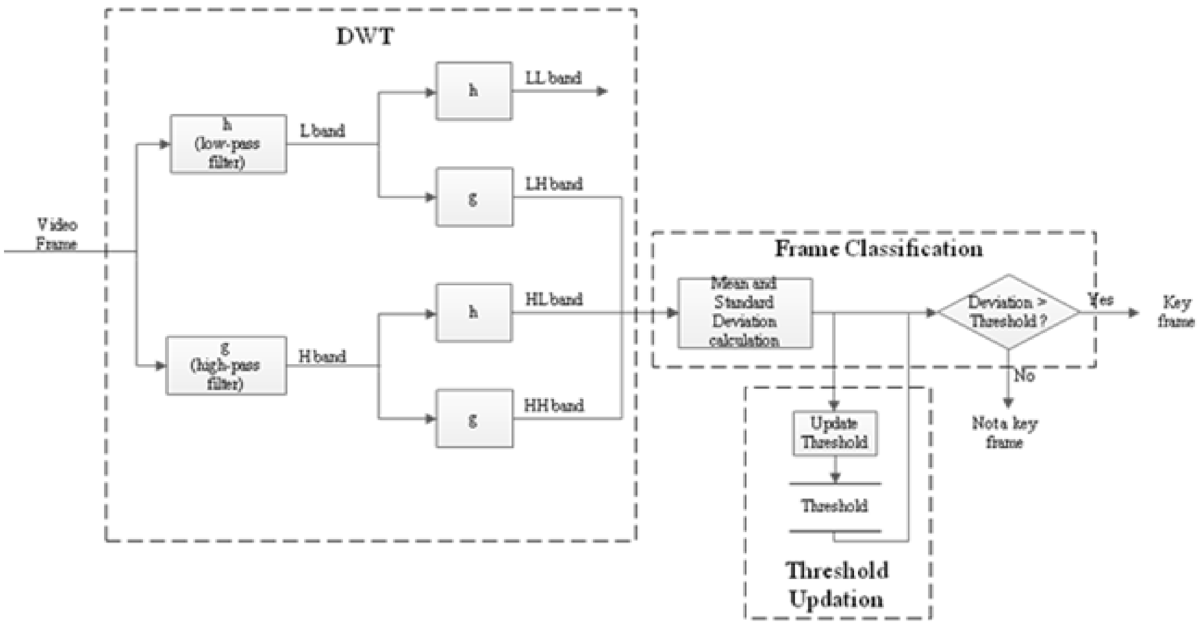

Fig. 3 Design of Key Frame Extraction

For the design illustrated above, Fig 4 gives the data flow diagram (DFD) for key frame extraction using DWT technique.

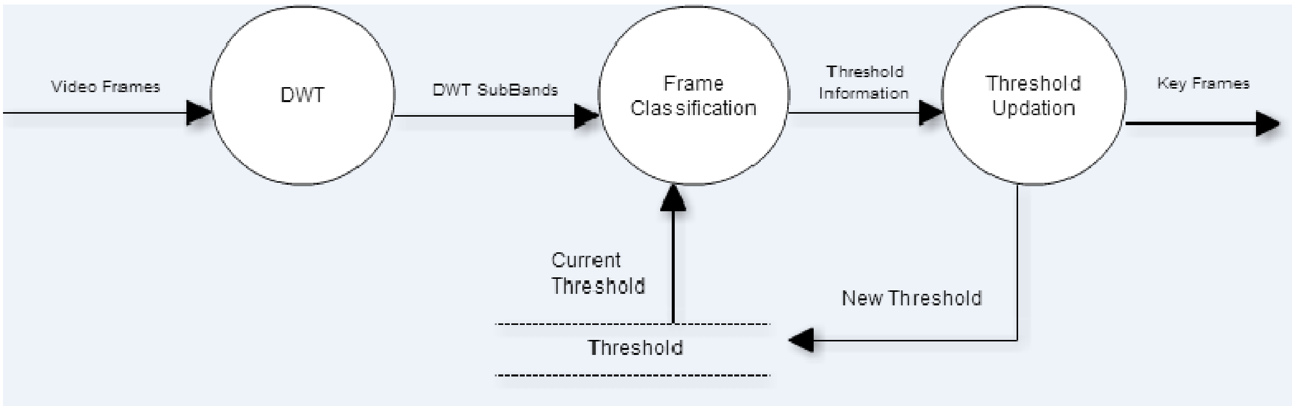

Fig. 4 Data Flow Diagram of Key Frame Extraction

Our proposed algorithm can be divided into four steps .In the first step, two successive frames are read and transformed with DWT to achieve four sub-bands, LL, HL, LH and HH. Within the four sub-bands, only three sub-bands, HL, LH and HH are used to detect key frame. For each sub-band, different value is estimated by subtracting detail component values of current and next frame. In the second step, Mean and Standard Deviation are computed from the difference values of each sub-band. In the step three, threshold value for each sub-band is calculated by adding the Mean and Standard Deviation. In final step, the threshold and difference value of each band are compared. If two difference values of any two sub-bands are over each related threshold, the last frame can be considered as a key frame [14].

The detailed algorithm can be illustrated with help of a flow chart which is given in Fig 5.

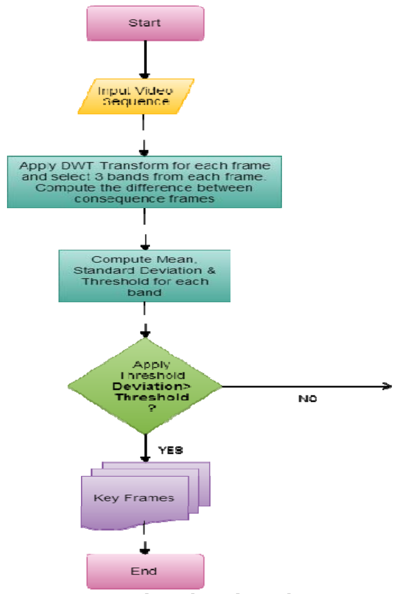

Fig. 5 Algorithm Flow Chart

III. Experimental Results

It was first proposed in [14], which the authors had implemented in Matlab but here in this paper I have implemented it using C language and I have made use of FFMPEG to encode and decode the video.

During test 1, video "Windows_7_Wildlife_Full_Version.mp4" was given as an input. A screen shot of Input video given in Fig. 6. Algorithm is implemented in C language and is run on Ubuntu OS using gc++ compiler. FFMPEG is made use to encode and decode the video. FFmpeg is the leading multimedia framework, able to decode, encode, transcode, mux, demux, stream, filter and play pretty much anything that humans and machines have created.

Fig. 6 Input video, *Windows_7_Wildlife_Full_Version.mp4*

It supports the most obscure ancient formats up to the cutting edge. No matter if they were designed by some standards committee, the community or a corporation. It contains libavcodec, libavutil, libavformat, libavfilter, libavdevice, libswscale and libswre sample which can be used by applications. FFMpeg is probably the most widely used Encoder/Decoder library. It helps you when are working in a project that involves video processing. As with any open source project, it has very limited documentation. An article by Martin Böhme (Using libavformat and libavcodec) and Stephen Dranger (FFmpeg and SDL Tutorial) were the only good articles available around. Apart from this, you might need to look intoOutputExample.c that comes with the FFMpeg source code. We would need to download MinGW and MSYS from ffmpeg site as a part of installation process.

As well as ffmpeg, ffserver, ffplay and ffprobe which can be used by end users for transcoding, streaming and playing. FFmpeg is a complete, cross platform command line tool capable of recording, converting and streaming digital audio and video in various formats. It can be used to do most of our multimedia tasks quickly and easily say, audio compression, audio/video format conversion, extract images from a video and a lot more [12,13].

The video “Windows_7_Wildlife_Full_Version.mp4” has 5015 frames for 2:48 minutes display time with many scene changes. The proposed method can retrieve 356 key frames for the video within 17 seconds. Fig. 7 gives few sample output key frames for input video.

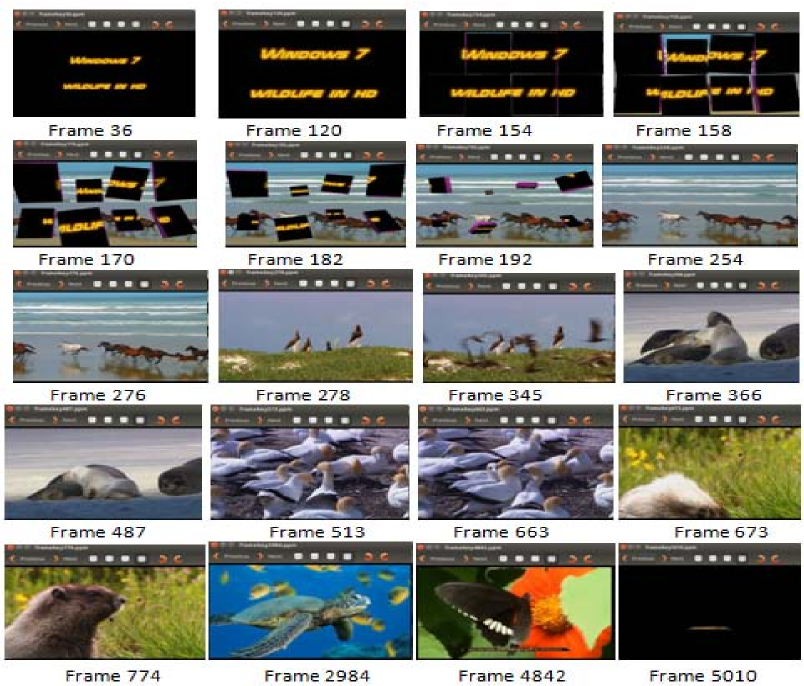

Fig. 7 Some output key frames of input video, *Windows_7_Wildlife_Full_Version.mp4*

During test 2, video "Modi_speech_15Aug2014.mp4" was given as an input. A screen shot of Input video given in Fig. 8. It has 7293 frames for 4:51 minutes display time with many scene changes. The proposed method can retrieve 773 key frames for the video within 98 seconds. Fig. 9 gives few sample output key frames for input video.

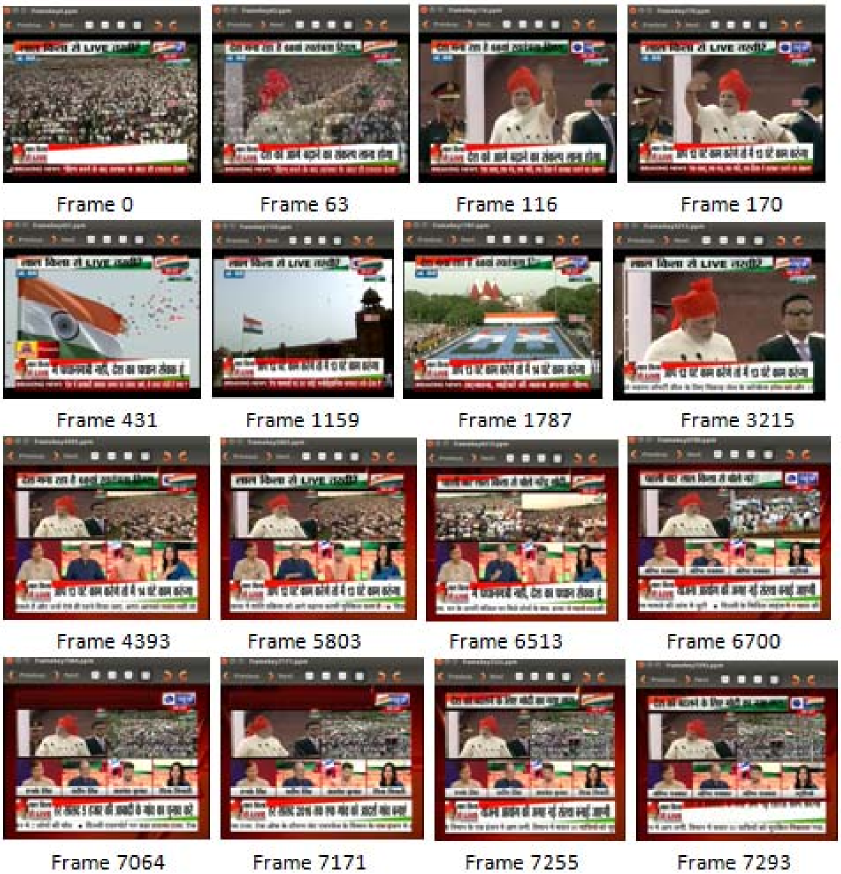

Fig. 9 Some output key frames of input video, *Modi_speech_15Aug2014.mp4*

IV. Conclusion

Digital images and videos are everywhere these days – in thousands of scientific (e.g., astronomical, bio-medical), consumer, industrial, and artistic applications. It is necessary to have automatic mechanisms of generating concise representation of the video content as a sequence of still or moving pictures i.e. video summary. The major task in video summarization is to segment the original video into shots and extract those video frames from the original video that would be the most informative and concise representation of the whole video. Such frames are referred as key frames. The extracted key frames should contain as much salient content of the shot as possible and avoid as much redundancy as possible. In this paper, DWT is considered for compressing the video. For each frame DWT is performed to get the DWT sub bands. Each of the LL, LH, HL and HH sub bands are obtained performing operations on the image and key frames are decided based upon the threshold values. It was first proposed in [14], which the authors had implemented in Matlab but here in this paper I have implemented it using C language and made use of FFMPEG to encode and decode the video.

References

1. https://www.coursera.org/course/digital
2. A.V.Kumthekar, Prof. J.K. Patil "Key frame extraction using color histogram method", International Journal of Scientific Research Engineering & Technology (IJSRET) Volume 2 Issue 4 pp 207-214 , ISSN 2278 – 0882, july-2013
3. Marian Kogler, Manfred del Fabro, Mathias Lux, Klaus Schoeffmann, Laszlo Boeszoermenyi "Global vs. Local Feature in Video Summarization: Experimental Results",in Proceedings of the 10th International Workshop of the Multimedia Metadata Community, Graz, Austria, 2009
4. Sachan Priyamvada Rajendra,Dr. Keshaveni N "A Survey of Automatic Video Summarization Techniques" International Journal of Electronics, Electrical and computational System(IJEECS), Volume 3,Issue 1 April 2014, ISSN 2348-117X
5. Truong, B. T. and Venkatesh, S. 2007. Video abstraction: A systematic review and classification. ACM Trans. Multimedia Comput. Commun. Appl. 3, 1, Article 3, Feb. 2007

6. RaviKansagara, DarshakThakore, MahaswetaJoshi , "A Study on Video Summarization Techniques", International Journal of Innovative Research in Computer and Communication Engineering Vol. 2, Issue 2, February 2014

7. Azra Nasreen,Dr. Shobha G "Key Frame Extraction from Videos - A Survey" International Journal of Computer Science & Communication Networks,Vol 3(3),194-198 ISSN:2249-5789

8. Amanjot Kaur, Jaspreet Kaur, "Comparision of Dct and Dwt of Image Compression Techniques", International Journal of Engineering Research and Development ISSN: 2278-067X, Volume 1, Issue 4 (June 2012), PP.49-52

9. Gonzalez, R. C. and Woods, R. E. and Eddins, S. L., Digital Image Processing UsingMATLAB, Prentice Hall, 2004.

10. Kamrul Hasan Talukder and Koichi Harada, "Development and Performance Analysis of an Adaptive and Scalable Image Compression Scheme with Wavelets", Published in the Proc. of ICICT, March 2007, BUET, Dhaka, Bangladesh, pp. 250-253, ISBN: 984-32-3394-8.

11. Chen, L., VLSI Design of Wavelet Transform: Analysis, Architecture and Design Examples Imp. College press, 2007

12. https://www.ffmpeg.org/about.html

13. http://rmn-explores.blogspot.in/2010/07/ffmpeg-building-on-windows_29.html

14. Khin Thandar Tint, Dr. Kyi Soe "Key Frame Extraction for Video Summarization Using DWT Wavelet Statistics", International Journal of Advanced Research in Computer Engineering & Technology (IJARCET) Volume 2, No 5, May 2013 ISSN: 2278 – 1323

AUTHOR(S) PROFILE

Chethan Sharma received his B.E degree in Computer Science and Engineering from NMIT, Bangalore and he is presently pursuing his M.Tech in Computer Science and Engineering from Christ University, Bangalore