ISSN: 2278 – 1323
International Journal of Advanced Research in Computer Engineering & Technology (IJARCET)
Volume 2, No 5, May 2013

# Key Frame Extraction for Video Summarization Using DWT Wavelet Statistics

**Khin Thandar Tint, Dr. Kyi Soe**

*Abstract*— **In the current information era, almost information are representing and processing in the forms of multimedia. Especially in video processing, numerous frames containing similar information are usually processed. This leads to unnecessary time consumptions, slow processing speed and complexity. Video summarization using key frames can facilitate to speed up video processing. In this paper, key frames extraction using wavelet statistics is discussed to use in video summarization. In extracting key frames, two consecutive frames are firstly DWT transformed and then the differences of the detail components of them are estimated. If different value of a consecutive pair is greater than threshold, the last frame of the pair is considered as a key frame. Experimental results are also discussed to represent the valid of the proposed method for video summarization.**

*Index Terms*— **video summarization, key frame extraction, DWT wavelet, differences of detail components, statistics.**

## I. INTRODUCTION

In the recent years, faster digital devices which can facilitate more storage space were rapidly developed. Hence, instead of representing information using texts and still images, audio, etc, the information were recorded and stored in video. By recording in video, information can be shown again and again without any doubt and confusing. Therefore, researchers are more and more interested in video processing and analysis.

Nowadays, information is record and used as video in various fields such as education, traffic control, environmental, research and so on. Actually, a video is composed of continuous still images called frames. For one second video, at least thirty frames are required. These frames are very similar on each others. However, when important information in the video are needed to display, all frames are not require. Only one frame that contains the important information is needed. This frame is usually called key frame. When we want to show complete information of a video in rapidly, only the key frames are needed to show instead of using all frames. Each key frame can provide and represent all necessary information of a video shot of the video. This display technique is usually called video summarization. It is also a compact representation of a video sequence. Hence, key frame representation and extraction is important feature of a video summarization.

The key frame extraction is also a part of video processing in which a frame containing all necessary information of a video shot is retrieved among other similar frames. The key frame extraction is not only used for video summarization, but also applied in other video processing such as video annotation, video transmission, video indexing, etc. Hence, researchers have been interested in the key frame extraction technology and tried to progress more and more in it.

The work area of key frame extraction is so wide and rich technology. Many techniques for key frame detection have been reported in so far. One of the possible approaches is the detection of video shots and the first frame of a shot is chosen as key frame [1]. Next interesting technique is the work of Seung et al. In this work, shot boundary was detected in low pass filtering in histogram space and key frame was selected by using adaptive temporal sampling. [2]. Furthermore, Mohamed and his colleagues published another histogram approach to extract key frame [3]. In his work, histograms of RGB channels of each consecutive frame were firstly estimated and found their differences. From the different values, threshold was calculated and compare with different value to choose key frame. Later, Hong *et al* report appearance based key frame detection in [4] in which key frame is extracted by using pixel wise, global histogram, local histogram, feature matching and Bags of Words (BoW). Another wavelet approach for key frame detection is issued by Gianluigi *et al* in [5]. In their work, key frame detection algorithm determined the complexity of video sequence in term of video content change. Three descriptor, colour histogram, wavelet and edge direction histogram were used to express the frame visual contents. In the last and one of the related works of our proposed method, the work of Khushboo *et al* [6] is wanted to present. This work is very sample but efficient. It is pixel wise processing in which the gradient differences of related pixels of consecutive frames is firstly estimated and then compare with threshold to determine key frame.

In this paper, key frame is detected by using wavelet detail coefficients. The wavelet detail coefficients can represent frame contents. Whenever frame contents are changed, the detail coefficients cannot be definitely the same. From the result, video shots and key frames can be easily detected. This paper is composed of five sections. First is introduction, second is background theory, third is proposed method, fourth is experimental result and conclusion is represented as last section.

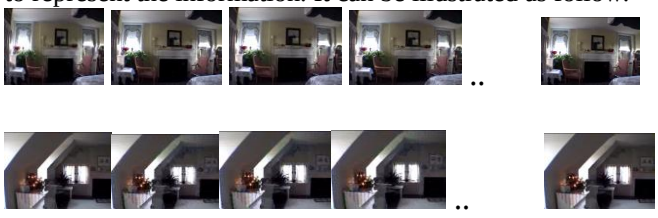
## II. BACKGROUND THEORY

In this section, the concepts of video frame and scene change, and about of DWT wavelet and its

detail frequency components are discussed as background of our proposed method.

### A. Concept of video frame and scene change

A video clip is composed of multiple frames. Actually, one video frame represents one still image so one scene of the video clip contain at least thirty frames depending on the frame rate of the video. The visual contents of the frames are not different too much but there are merely nibble changes to be represented video motion in them. Hence, the thirty frames or one scene can be taken only single information of the video. Whenever we wanted to display other information, the scenes are needed to change. Hence, when we want to show only the information of the video as a summarization, we do not need to show all frames of a scene and merely one frame is enough to represent the information. It can be illustrated as follow.



(a) (b)

Fig. 1 Illustration of (a) video frames of a scene and (b) selected frame that represents information of the scene

In the figure 1, there are two scenes containing multiple frames. As scenes are different, visual contents in the frames of each scene are also different. The visual contents can be represented by two approaches, spatial (pixel) and frequency (colour). In the spatial approaches, the visual contents of a frame or an image are composed of pixels with different gray level values. Hence, if some of related pixel values of two frames are different, the visual contents of the two frames are also different. Similarly in frequency, if frequency components of two images are different, the images are exactly different. By this way, scene changes can be detected by finding the difference over a threshold.

### B. DWT wavelet and its frequency components

Wavelet processing is one of the frequency domain based processing. When an image is performed by DWT wavelet, the following four sub band images with different properties are achieved.

As shown in figure 2, in the four sub-band transformed images, LL corresponds to a smooth version of original image. HL, LH and HH are the three coefficients of details. Hence, the obviously change in original image can cause the changes in coefficient values of the three sub bands [7].

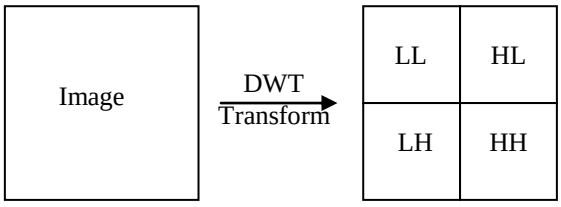



Fig. 2 Illustration of DWT wavelet transformation of an image

According to the above wavelet expression, the three sub bands coefficient change is one of the possible approaches to detect scene change and extraction of key frame. In our proposed method, we use the concept of the wavelet as discussed above.

## III. Proposed Method

Our proposed method is key frame extraction from different scenes of a video clip. Each key frame can represent each related scene and also entirely contains all important information of the scene. After the key frame extraction, the key frames are intended to use in video summarization, feature extraction and other processing so key frame extraction algorithm should not be very complex and time consuming. Our objectives for the proposed algorithm are specially focused on speed and precise.

In our proposed method, for the detection of key frame from a scene, we have to find the obvious difference value between two successive frames. In a video stream, each video frame is a slightly variation with previous one. However, whenever scenes are changed, visual contents and objects are obviously different between current frame and next one. Hence, we use the difference as a key for the key frame detection.

Several approaches such as sampling based, clustering based, shot-based, etc have been proposed for the key frame extraction. However, unlike previous works, our proposed method is considered and based on DWT wavelet according to the above wavelet discussion in section II. Whenever visual contents are changed in original image, the frequency components of transformed image are also changed. When scenes are changed, the visual contents of current and later frames are not the same. Hence, the frequency components of the frames are exactly different. By using the difference, we can detect the key frame of next scene. In our proposed method, the reason for using DWT wavelet is that sub-band frequency components can be easily distinguished by each group in it.

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. The algorithm of the proposed method can be seen as follow.

Input: Video V, contains N frames
Output: Key frame for input video

---

Algorithm of Key Frame Extraction

**Begin**
Read Video V;

**Step 1**

For loop of each video frame k = 1 to N

1. Read video frame $V_k$ and $V_{k+1}$
2. Eliminate RGB frame to Gray image
   $G_k$ = gray image of $V_k$
   $G_{k+1}$ = gray image of $V_{k+1}$
3. Gray image is transformed by level 1 of DWT to obtain four channel sub bands
   cH1 = HL band of $G_k$
   cV1 = LH band of $G_k$
   cD1 = HH band of $G_k$
   cH2 = HL band of $G_{k+1}$
   cV2 = LH band of $G_{k+1}$
   cD2 = HH band of $G_{k+1}$
4. Find difference total of each band

$$D1(k) = \sum_{i-1}^{m}\sum_{j=1}^{n}\left(cH2(i,j) - cH1(i,j)\right)$$

$$D2(k) = \sum_{i-1}^{m}\sum_{j=1}^{n}\left(cV2(i,j) - cV1(i,j)\right)$$

$$D3(k) = \sum_{i-1}^{m}\sum_{j=1}^{n}\left(cD2(i,j) - cD1(i,j)\right)$$

Where, *m* and *n* are numbers of row and column. *D1*, *D2* and *D3* are difference of *cH*, *cV* and *cD* bands.

End of for loop.

**Step 2**

Compute Mean & Standard Deviation

1. Mean (M) calculation

$$M1 = \frac{\sum_{i=1}^{N-1}\left(D1(i)\right)}{N-1}$$

$$M2 = \frac{\sum_{i=1}^{N-1}\left(D2(i)\right)}{N-1}$$

$$M3 = \frac{\sum_{i=1}^{N-1}\left(D3(i)\right)}{N-1}$$

2. Standard Deviation (STD)

$$STD1 = \sqrt{\frac{\sum_{i=1}^{N-1}\left(D1(i)-M1\right)^2}{N-1}}$$

$$STD2 = \sqrt{\frac{\sum_{i=1}^{N-1}\left(D2(i)-M2\right)^2}{N-1}}$$

$$STD3 = \sqrt{\frac{\sum_{i=1}^{N-1}\left(D3(i)-M3\right)^2}{N-1}}$$

**Step 3**

Estimation of Threshold
Th1 = M1 + αSTD1
Th2 = M2 + αSTD2
Th3 = M3 + αSTD3
Where, α is a constant.

**Step 4**

For k =1 to N-1
If ((D1(k)>th1 AND D2(k)>Th2) OR (D2(k)>th2 AND D3(k)>Th3) OR (D1(k)>th1 AND D3(k)>Th3) )
Write $V_{k+1}$as output key frame
End of for loop

**End** of algorithm

The above proposed algorithm is not complex too much and can easily and quickly detect key frames of a video clip. The results of the proposed algorithm can be seen in next section.

## IV. Experimental Results

In this section, three different video clips, "*Housetour.avi*", "*Telugu.avi*" and "*farm.mp4*" are used as inputs to the proposed system. Some videos are downloaded from You Tube. In our proposed system, audio parts of the videos are neglected. Only visual content changes in the videos are mainly focused to detect key frame. The proposed algorithm is implemented in Matlab 2012a.


Fig. 3 Some frames of input video, *Housetour.avi*

. In the above figure 3, some frames of input video clip, 'housetour.avi' are shown. The video contains 663 frames. When the differences between two consecutive frames are estimated, frame 1 and 2, frame 2 and 3 are not obviously different in all sub bands but frame 3 and 4 are distinctly different with the value, D1= 43.00 which is over the threshold, Th1= 40.6204 and D3=19.75 which is over Th3=10.3249. Hence, frame 4 is remarked and saved as a key frame. Similarly, other frames of the video are compared with related threshold and detected to produce key frames. For the video clip, the values contains in the following table are used to detect key frames.

TABLE I

PARAMETERS AND VALUES USED IN DETECTION OF KEY FRAME

| Variable | Symbol | Value |
|---|---|---|
| Alpha | α | 0.55 |
| Mean | M1 | -0.2315 |
| | M2 | 0.2376 |
| | M3 | 0.0520 |
| Standard Deviation | STD1 | 74.2762 |
| | STD2 | 515.8240 |
| | STD3 | 18.6779 |
| Threshold | Th1 | 40.6204 |
| | Th2 | 283.9408 |
| | Th3 | 10.3249 |

In detecting key frames using the parameters and values in table 1, totally 65 key frames are detected for the video, *Housetour.avi*. In the detecting the key frames, the constant, alpha (α) is really noticeable and traded off because exceed key frames will be produced if smaller alpha value is used. On the otherwise, if the alpha is so large, key frames cannot be produced in every scene changes. Some key frames produced by the proposed method for the *Housetour.avi* are summarized as followed.

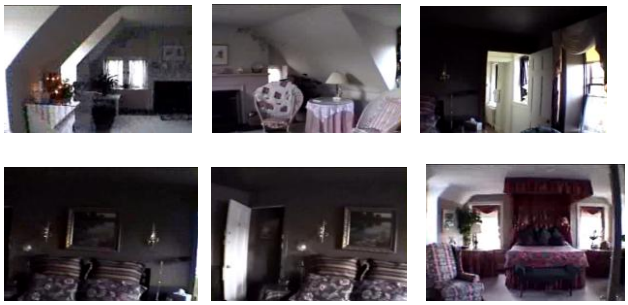


Fig. 4 Some output key frames of input video, *Housetour.avi*

As a next experiment, key frames are extracted from farm.mp4 which has more than 3000 frames for 2 minutes. In this video, each scene has about 60 frames and there are round about 50 scenes. The proposed method can retrieve more than 60 keys frames for the video. Some scenes take longer display time so they take more frames. For the scene, key frames are extracted more than one. Some frames containing in the video, farm.mp4 and its result key frames are described as follow.


Fig. 5 Some frames of input video, *farm.mp4*

Fig. 6 Some output key frames of input video, *farm.mp4*

Finally, we test the proposed method with a trailer video, "*Telugu.avi*" which is downloaded from internet. It has 501frames with so many scene changes. It takes totally one minute and two second display time. Our proposed method can retrieve 20 key frames for the video within 33 seconds. The output key frames are as followed.

Fig.. 7 Some output key frames of input video, *Telugu.avi*

## V. CONCLUSION

In this paper, we have presented key frame generation algorithm using DWT wavelet. The objective of the method is to generate key frames with faster speed for other video processing such as video summarization, feature extraction, etc. According to experimental results, key frames can be precisely generated in almost of scenes. Furthermore, we achieved that processing speed of the system was relatively fast. Hence, the proposed method can hold our objectives. Video processing and analysis is one of the interesting technologies. As our future works, we will continue our researches on this field, especially in feature extraction and pattern recognitions from a video.

## REFERENCES

[1] Tonomura Y., Akutsu A., Otsugi K., and Sadakata T "VideoMAP and VideoSpaceIcon: Tools for automatizing video content" *Proc. ACM INTERCHI 93*, pp 131-141, 1993.

[2] Seung Hoon Han, Kuk Jin Yoon and In So Kweon "A new technique for shot detection and key frame selection in histogram space" *Workshop on Image Processing and Image Understanding*, 2000, pp 305-310, 2000.

[3] Mohamed Ahmed, Ahmed Karmouch and Suhayya Abu-Hakima "Key frame extraction and indexing for multimedia database", *Image and Vision Interface 99*, pp. 19–23, May 1999, 99, Trois-Rivières, Canada.

[4] Hong Zhang, Bo Li and Dan Yang, "Key frame detection for appearance based visual SLAM," *IEEE/RSJ International Conference*, pp 2071-2076, October 2010, Taipei, Taiwan.

[5] Gianluigi Ciocca and Raimondo Schettini, "Dynamic key frame extraction for video summarization," Proc. ACM Multimedia, pp 257-262, November 1999.

[6] Khushboo and M. B. Chandak, "Key frame extraction methodology for video annotation," *International Journal of Computer Engineering and Technology*, vol. 4, pp 221-228, March 2013.

[7] Tinku Acharya and Ajoy K. Ray, "Image Processing Principle and Application," John Wiley & Sons, Inc., Hoboken, New Jersey, Canada, 2005.