International Journal of Science and Research (IJSR)
ISSN (Online): 2319-7064
Impact Factor (2012): 3.358

Comparison of Video Shot Detection and Video Summarization Techniques

Vijay Kumar D[1], Santhosh Kumar K.L[2], Jharna Majumdar[3]

[1]M.Tech Scholar, Department of CSE, Nitte Meenakshi Institute of Technology, Bangalore, India

[2]Assistant Professor, Department of CSE (PG), Nitte Meenakshi Institute of Technology, Bangalore, India

[3]Dean R&D, Professor and Head CSE (PG), Nitte Meenakshi Institute of Technology, Bangalore, India

Abstract: ***The Process of shot boundary detection is a fundamental requirement in automatic video indexing, editing and archiving. Many algorithms have been proposed for detecting video shot boundaries and classifying shot and shot transition types. This paper presents a comparison of several new shot boundary detection and classification techniques and their variations including Histograms, Discrete wavelet transform, Haar wavelet based video shot detection and VGRAPH Methods. The performance and ease of selecting good thresholds for these algorithms are evaluated based on a wide variety of video sequences. Threshold selection requires a trade-off between recall and precision that must be guided by the target application.***

Keywords: DWT, Haar Wavelet, VGRAPH, Video Shot Detection, Video Summarization.

1. Introduction

With the widespread usage of video in computer systems and networks, automatic video content summarization techniques are becoming a necessity. A concise and informative video summary will enable the user to quickly figure out general content of a video and help him/her to decide if it is worthwhile to watch through whole sequence. For video content retrieval, a video summary will dramatically save the user's time and efforts to spot the desired videos from large volume of video collections.

One of the major preludes to video summarization is video shot detection (VSD). Recently, several approaches have been proposed. These methods segment the video into camera shots and extract fixed number of key frames from each shot. The major techniques that have been used for shot boundary detection are DWT [11] co-efficient band differences, Harr [12] Wavelet co-efficient band differences, Histogram difference and VGRAPH [14].

2. Objective

There has been a good deal of research on using computers to do automatic content extraction of videos. Here we focused on cut detection, and more recent work has focused on detecting gradual transitions. The major techniques that have been used for shot boundary detection are DWT, Haar, and VGRAPH. The only reported comparisons of these techniques applied the tested methods to a small number of short test sequences and sometimes tuned the methods to work well on those sequences.

3. Proposed Algorithms for Video Shot Detection and Video Summarization Methods

3.1 Dwt Based Video Shot Detection and Video Summarization

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.

The visual contents can be represented by two approaches, spatial (pixel) and frequency (color). 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.

Wavelet processing is one of the frequency domain based processing. When an image is performed by DWT wavelet [11], the following four sub band images with different properties are achieved. As shown in figure 1, 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.

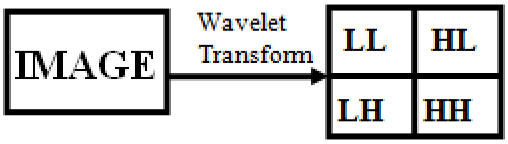

Figure 1: Wavelet Transform image

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.

Steps for Video shot detection by DWT Wavelet.

- Step1: Read all frames from given input video and eliminate RGB frame to grey.
- Step2: Grey frame is transformed by level 1 of DWT to obtain four sub bands co-efficient as shown in fig1.
- Step3: Calculate Difference of three bands except the LL band in four sub bands.
- Step4: Calculate mean, Standard Deviation and Estimate the Threshold
- Step5: Apply Threshold to identify shots and non – shots.

3.2 Harr Wavelet Based Video Shot Detection:

Haar functions are used since 1910[12]. They were introduced by Hungarian mathematician Alfred Haar .These transforms have been applied, for instance, to spectral techniques for multiple valued logic, image coding, edge extraction, etc. A complete orthogonal system of functions in $L^p [0, 1]$, $p \in 2 [0, \infty]$ which take values from the set $\{0, 2^j : j \in N\}$ was defined by Haar. This system of functions has property that each function continuous on interval (0, 1) may be represented by a uniformly and convergent series in terms of elements of this system. The Haar definition is shown below

$$\textbf{Haar}(0, t) = 1, \text{ for } t \in [0, 1];$$

$$\textbf{Haar}(1, t) = \begin{cases} 1, \; for\; t \in \left[0, \frac{1}{2}\right] \\ -1, \; for\; t \in \left[\frac{1}{2}, 1\right] \end{cases}$$

Discrete Haar functions may be defined as functions determined by sampling the Haar functions at 2^n points. These functions can be conveniently represented by means of matrix form. The Haar matrices H(n) are considered in the natural and sequence ordering which differ in the ordering of rows. Each row of the matrix H(n) includes the discrete Haar sequence haar(w, t). In this notation, index w identifies the number of the Haar function and index t the discrete point of the function determination interval. In this case, the Haar matrix of any dimension may be obtained by the following recurrence relation: $X=2^{(n-1)/2}$

$$H(n)=\begin{bmatrix} H(n-1) \otimes [1 \quad 1] \\ X\, I(n-1) \otimes [1 \quad -1] \end{bmatrix}, \; H(0)=1$$

The Haar matrix is non–symmetric and its elements are 1, −1 or 0, multiplied by powers of √2. The discrete, orthogonal Haar functions, obtained from the above formula, are defined on [0, 1] interval. i.e fig 2 Shows Haar wavelet transformed Image.

Steps for Video shot detection by Haar Wavelet.

- Step1: Read all frames from given input video and eliminate RGB frame to grey.
- Step2: Grey frame is transformed by level 1 of Haar to obtain four sub bands co-efficient as shown in fig 2.
- Step3: Calculate Difference of three bands except the LL band in four sub bands.
- Step4: Calculate mean, Standard Deviation and Estimate the Threshold
- Step5: Apply Threshold to identify shots and non- shots

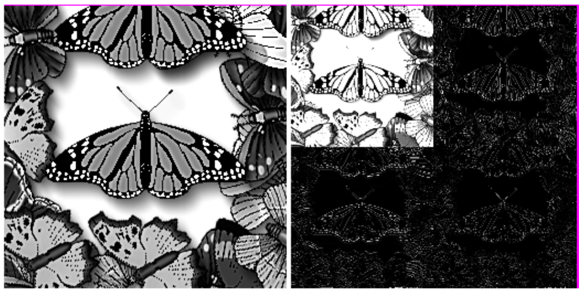

Figure 2: Wavelet Transform image

3.3 VGRAPH

An Effective Approach for Generating Static video Summarization. Steps of VGRAPH Approach [14]:

- Step1: The Original Video is Pre- Sampled. The target of the pre sampling step is to reduce the number of frames to be processed. Selecting suitable sampling rate is very important; as low sampling rate leads to poor video summaries; while large sampling rate shortens the video summary. VGRAPH approach is selected to be one frame per second.
- Step 2: Pre sampled video is segmented into shots using the color features. For Video temporal segmentation step, a simple shot boundary detection method is used
- Step 3: The noise frames are eliminated and representative frame is selected from each shot.
- Step 4: Key frames are extracted using nearest neighbor graph which is built from texture features extracted from shot representative frames.

4. Experimental Results

We have examined four popular techniques of shot boundary detection and results are compared. All the methods implemented in this paper are tested on a wide range of thresholds. For each combination of algorithm and threshold set, we measure the number of shot boundaries that were correctly detected, number of false positive and number of missed boundaries. It is a common way to choose Recall and Precision as the appropriate evaluation criteria.

Recall = (correct) / (correct + missed)

Precision = (correct) / (correct + false positive)

In many applications, a treadoff must be made between the recall and precision. It may or may not be acceptable to retrieve extra shot boundaries that would otherwise be missed at the expense of retrieving of each method,

Sample Input video 1

Figure 4: Input Video FootBall.avi

5. Results

1. DWT based Video Shot Detection & Summarization.

Figure 5: DWT Output Key Frames for Video

2. HAAR based Video Shot Detection & Summarization

Figure 6: HAAR Output Key Frames for Video

3. VGRAPH based Video Shot Detection & Summarization

Figure 8: VGRAPH Output Key Frames for Video

Table1: Video test data

Method	*Video Name*	*Total Frames*	*Actual Cut*	*Shot Detected*	*Shot Miss*	*False Detect*
DWT	1.FootBall	1711	9	9	0	0
	2.Park	617	5	4	1	0
	3.Wild	742	7	5	2	2
	4. Cut3	610	7	6	1	0
Haar	1.FootBall	1711	9	8	2	1
	2.Park	617	5	5	1	1
	3.Wild	742	7	4	3	0
	4. Cut3	610	7	8	0	1
VGRAPH	1.FootBall	1711	9	11	1	3
	2.Park	617	5	5	1	1
	3.Wild	742	7	11	1	5
	4. Cut3	610	7	10	1	4

Table 2: Recall and Precision for test data

Method	Video Name	Recall	Precision	Computation Time
DWT	1.FootBal	1.00	1.00	213.87
	2.Park	0.80	1.00	19.12
	3.Wild	0.71	1.00	69.7
	4. Cut3	0.85	1.00	67.49
Haar	1.FootBal	0.80	0.88	53.01
	2.Park	0.83	0.83	19.71
	3.Wild	0.57	1.00	22.97
	4. Cut3	1.00	0.88	57.24
VGRAPH	1.FootBall	0.91	0.78	25.65
	2.Park	0.83	0.83	9.85
	3.Wild	0.91	0.66	11.11
	4. Cut3	0.90	0.69	28.62

6. Conclusion

A combination of these three wavelet band features might produce better results than either the region histogram or the running histogram algorithms. In general, the simpler algorithms outperformed the more complicated algorithms. These complicated algorithms were sensitive to the threshold settings and ‘‘hidden’’ parameters not specified in the literature.

Dwt Based Shot Detection method takes more time to compute the co efficient bands, Haar Wavelet method takes less time to compute the bands and VGRAPH method which will take the advantage of both Wavelet and which uses the color feature and the Histogram to detect the shots. Based on the shots Key frames are Summarized.

References

[1] H. J. Zhang, A. Kankanhalli, and S. W. Smoliar, ‘‘Automatic partitioning of full-motion video,’’ Multimedia Systems **1(1)**, 10–28 (1993).

[2] A. Nagasaka and Y. Tanaka, ‘‘Automatic video indexing and fullvideo search for object appearances,’’ in Visual Database Systems II, E. Knuth and L. Wegner, Eds., pp. 113–127, Elsevier Science Publishers (1992).

[3] R. Zabih, J. Miller, and K. Mai, ‘‘A feature-based algorithm for detecting and classifying scene breaks,’’

Proc. ACM Multimedia 95, pp. 189–200, San Francisco, CA (1995).
[4] B. Shahraray, ''Scene change detection and content-based sampling of video sequences,'' in Digital Video Compression: Algorithms and Technologies, Proc. SPIE **2419**, 2–13 (1995).
[5] A. Hampapur, R. Jain, and T. Weymouth, ''Digital video segmentation,'' Proc. ACM Multimedia 94, pp. 357–364, San Francisco, CA (1994).
[6] R. Kasturi and R. Jain, ''Dynamic vision,'' in Computer Vision: Principles, R. Kasturi and R. Jain, Eds., IEEE Computer Society Press, Washington (1991).
[7] H. Ueda, T. Miyatake, and S. Yoshizawa, ''IMPACT: an interactive natural-motion-picture dedicated multimedia authoring system,'' Proc. CHI. 1991, pp. 343–350 ACM, New York (1991).
[8] D. Swanberg , C. F. Shu, and R. Jain, ''Knowledge guided parsing and retrieval in video databases,'' in Storage and Retrieval for Image and Video Databases, Proc. SPIE **1908**, 173–187 (1993).
[9] T. D. C. Little, G. Ahanger, R. J. Folz, J. F. Gibbon, F. W. Reeve, D. H. Schelleng, and D. Venkatesh, ''A digital on-demand video service supporting content-based queries,'' Proc. ACM Multimedia 93, pp.427–436, Anaheim, CA (1993).
[10] F. Arman, A. Hsu, and M-Y. Chiu, ''Image processing on encoded video sequences,'' Multimedia Systems **1(**5), 211–219 (1994).
[11] Khin Thandar Tint, Dr. Kyi Soe "Key Frame Extraction for Video Summarization Using DWT Wavelet Statistics" (IJARCET) Volume 2, No 5, May 2013.
[12] Piotr Porwik, Agnieszka Lisowska "The Haar–Wavelet Transform in Digital Image Processing Its Status and Achievements" Machine GRAP HI CS & VI S I O N vol 13, n o. 1/2, 2004, p p .79-98.
[13] Brojeshwar Bhowmick and Debaleena Chattopadhyay "Shot Boundary Detection Using Texture Feature based On Co-occurrence Matrices" 978-1-4244-3604-0/09 IMPACT- 2009.
[14] Karim M.Mahmoud, Nagia M. Ghanem And Mohamed A. Ismail," VGRAPH:An Effective Approach for Generating Static Video Summaries " DOI: 10.1109/iccvw.2013.111.

Author Profile

Vijay Kumar D Completed B.E in Computer Science from Sir M Visvesvaraya Institute of Technology, Banglore in 2010 and currently pursuing M.Tech in Nitte Meenakshi Institute of Technology, Bangalore. His areas of interest are Image and Video Processing, Video Shot Detection.

Santhosh Kumar K L received B.E and M.Tech in Computer Science and Engineering from Visvesvaraya Technological University. He is currently working as Assistant Professor in the Department of Computer Science & Engg., Nitte Meenakshi Institute of Technology, Bangalore. His areas of interest are Image and Video processing, Computer Vision and Pattern recognition.

Dr. Jharna Majumdar currently working as Dean R & D and Professor and Head of Computer Science and Engineering (PG) at the Nitte Meenakshi Institute of Technology, Bangalore. Prior to this Dr. Majumdar served Aeronautical Development Establishment, Defence Research and Development Organization (DRDO), Ministry of Defence, Govt. of India as Research Scientist and Head of Aerial Image Exploitation Division, Bangalore. Dr. Majumdar has 40 years of experience in R & D and Academics in the country and abroad. She has published large number of papers in National, International Journals and Conferences. Her Project with team of students from 7 engineering colleges and ISRO for building the first smallest satellite in India (a PICO satellite of weight less than 1 kg) had taken off to the orbit successfully by mid May 2010. Nitte Amateur Satellite Tracking Centre (NASTRAC) developed by a team of students from NMIT, Bangalore, under her guidance is the first Tracking Station of Small Satellites developed in the country.

The research team of robotics under her guidance has developed a Robot with Innovative Vision System and installed at the Birla Science Centre, Hyderabad as the first Robotics Exhibit in Indian Museums. Dr. Majumdar has 3 patents with students of NMIT, Bangalore in last 4 years.