



Analysis and Verification of Video Summarization using Shot Boundary Detection

Sowmya R¹, Dr. Rajashree Shettar²

¹Department of Computer Science and Engineering,
R V College of Engineering, Bangalore
Karnataka, India

²Professor, Department of Computer Science and Engineering,
R V College of Engineering, Bangalore
Karnataka, India

Abstract: The digital storage in this technology world has played a major role in all most all of our routine applications. It contains a large amount of data including videos, extracted features, alerts, statistics etc. Designing systems to manage this extensive data and make it easily accessible for query and search is a very challenging and potentially rewarding problem. However, the vast majority of research in video indexing has taken place in the field of multimedia, in particular for authored or produced video such as news or movies, and spontaneous and broadcast video such as sporting events. This paper mainly focuses on the analysis of the video using the shot boundary detection methods. Shot boundary detection is the fundamental step in the content based video analysis. It is even a major research issue, since this has been used as an important parameter in the video retrieval process. The analysis in this paper contains two different methods, 1. Block based Histogram difference and 2. Block based Euclidean distance difference for varying block sizes. The results show that as the block size increases the Histogram method performs better. But, block size does not impact performance in Euclidean distance method. Overall results are better in Euclidean distance method than the Histogram method.

Keywords: Shot boundaries, Video transitions, Video content analysis, Block based Histogram difference, Block based Euclidean distance difference.

I. Introduction

Videos, one of the most important media, have been widely used in many applications for information recording. For the increasing requirement of information mining, video analysis has drawn more and more attentions with the purpose to realize effective video content access. The increasing availability of digital video lacks from the increase in its accessibility. This is because of the structure of video data, which is unsuitable for traditional forms of data access, indexing, browsing, and retrieval. Traditional forms of data retrieval are either text based or based on the query-by-example. If we want to find a clip of interest, we have to sequentially browse through the video. This is an extremely time consuming, tedious and labor-intensive process. This directly arises a need of new technology or a tool for effective, efficient indexing and retrieval of the video data. This area plays an extensive role in the research area as well. Video structure analysis is a main and basic foundation for the video indexing and retrieval process. Generally, videos are structured according to a descending hierarchy of video clips, scenes, shots, and frames. Video structure analysis aims at segmenting a video into a number of structural elements that have semantic contents, including shot boundary detection, key frame extraction, and scene segmentation. Video shot boundary detection, which segments a video by detecting boundaries between camera shots, is usually the first and important step for content-based video retrieval. Digital video data can be processed with an objective of extracting the information about the content, which has embodied in the video data. The algorithms which has been developed to perform this purpose is referred as "video content analysis" algorithms and server the basic for developing tools, that helps to understand the events and objects within that particular scene of the video. As mentioned earlier, the first step of content based video retrieval is shot boundary detection. A shot is defined as a sequence of frames taken by a single camera with no major changes in the visual content.

Shot boundaries can be broadly classified into two types: abrupt transition and gradual transitions. Abrupt transition is instantaneous transition from one shot to the subsequent shot such as abrupt cuts. Gradual transition occurs over multiple frames, such as fade, dissolve and wipe. The detection of the transitions helps to evaluate

the performance of the shot boundary detection methods. This paper implemented abrupt cut and fade in/out methods to evaluate the performance of the shot boundaries detected. The abrupt cut is detected using Mean Square Error (MSE) approach [7]. The fade in/out transitions are detected using joint entropy approach [8].

I. Literature Survey and Related work

There are many research papers with respect to video content based analysis, where in Shot boundary Detection (SBD) is the basic and fundamental step for video analysis. Shot is a process of segmentation of video where boundaries are detected in camera shots. Hence, SBD is a popular area in the video content analysis and has been studied for a long time. Research has resulted in a variety of algorithms. In this section, briefly the SBD work in the literature is reviewed. Authors in [1], have proposed an approach for detection of Shot boundaries using block based histogram difference and extracted the key frames based on the shot boundaries. The performance has evaluated by identifying the video transitions such as cut, wipe and fade in/out. One of the important observations in this paper is the block size. Block size impacts the performance of the histogram based shot boundary detection. In [2], authors have implemented various automatic shot boundary detection algorithms in two different respects: newer algorithms designed explicitly to detect more complex editing operations such as fades and dissolves are taken into account; on the other hand, besides the algorithms' ability to detect edits as such, also their ability to classify the types of edits and locate their boundaries are examined. In [3] authors have presented a novel algorithm for simultaneously detecting shot boundaries and extracting key frames from video sequences or streams in real-time. The features extracted from the video are represented as a feature vector and arranged in a matrix. This approach does not provide a significant accuracy and efficiency. In [4], authors proposed an algorithm for the shot boundary detection using Discrete Wavelength Transform (DWT) and support vector machine (SVM). The results evaluated for both the methods had good accuracy and computational speed. But the methodology is very complex for implementation. In [5] authors presented a hardware/software code design solution to the shot boundary detection using SVM as a recognition engine. This has been implemented on the Field Programmable Gate arrays [FPGA]. But this is suitable only for the embedded system designer to adopt into their designs. In [6] authors presented details based on the different video structural analysis for the semantic contents like shot boundary detection, key frame extraction and scene segmentation. The different methods with respect to each of the semantic content have been discussed with advantages and limitations as well. The main motivating factor for the video summarization presented in this paper is to build an efficient foundation for the video indexing and retrieval mechanism. The performance of video summarization plays a real role in achieving performance in video data retrieval. Since, the video summarization removes the redundant data and makes an abstract summary of the contents; this is exhibited to users in a readable form for browsing. Video summarization directly impacts on video retrieval, by providing browsing of retrieved videos faster, especially for the large sized videos.

II. Implementation Strategy

The shots are basic building blocks of the video, in which this is the first step in the video summarization mechanism. The shots are the continuous frames from start to the end of recording in a camera. The implementation in this paper consists of below three modules.

1. Block based Histogram difference to calculate the Shot Boundary Detection (SBD).
2. Block based Euclidean distance difference to calculate the Shot Boundary Detection (SBD).
3. Video transition extraction such as cuts, fade in/out.

1. Block based histogram difference for SBD:

Step 1: Input a video, divide it in to frames.

Step 2: Separate R, G and B component of each frame of the video.

Step 3: Divide frame into number of blocks B [i, j]. The size of the block is 9, 25, 64 and 100.

Step 4: Compute the histogram matching difference between corresponding blocks in consecutive frames based on following equation (3.1):

$$Dg(k, k + 1, i, j) = \sum_{i=0}^{L-1} \frac{[H(i, j, k) - H(i, j, k + 1)]^2}{H(i, j, k)} \quad (3.1)$$

Where L = number of gray in the image,

B (i,j) = block at (i,j), blocks with m rows and n columns

H (i,j,k) = histogram of blocks at (i,j) in the kth frame,

H (i, j, k+1) = histogram of blocks at (i, j) in the k+1th frame,

$Dg(k, k + 1, i, j)$ = histogram difference between k^{th} and $k+1^{\text{th}}$ consecutive frames of an image.

Step 5: Compute histogram difference between two consecutive frames based on the following equation 3.2:

$$D(k, k + 1) = \sum_{i=1}^m \sum_{j=1}^n w_{ij} [D_B(k, k + 1, i, j)] \quad (3.2)$$

Where w_{ij} = weight of the block at (i,j)

Step 6: Compute the Mean Deviation (MD) of histogram difference over a whole video sequence.

Step 7: Compute the Standard Deviation (STD) of histogram difference over a whole video sequence.

Step 8: Compute Threshold $T = MD + a \cdot STD$,

Where a = threshold factor whose values are 0, 0.1, 0.2, 0.3 ... 1.

Step 9: If $D(k, k+1) = T$, Then k is the end frame of previous shot and $k+1$ is start frame of next shot.

2. Block based Euclidean difference for SBD

Step 1: Input a video, divide it in to frames.

Step 2: Separate R, G and B component of each frame of the video.

Step 3: Divide frame into number of blocks $B[i,j]$. Size of block is 9, 25, 64 and 100.

Step 4: Compute the Euclidean difference between corresponding blocks in consecutive frames.

Step 5: Compute Euclidean difference between two consecutive frames.

Step 6: Compute the Mean (MD) of Euclidean difference over a whole video sequence.

Step 7: Compute the Standard Deviation (STD) of Euclidean difference over a whole video sequence.

Step 8: Compute Threshold $T = MD + a \cdot STD$, where a = threshold factor, whose values are 0, 0.1, 0.2, 0.3 ... 1.

Step 9: If $D(k, k+1) = T$, Then k is the end frame of previous shot and $k+1$ is start frame of next shot.

3. Video transition extractor

Step 1: Calculate the abrupt cut using the Mean Square Error method [7].

Step 2: Calculate the fade in/out using the Joint Entropy approach [8].

III. Experimental Results and Discussion

Shot boundary detection algorithms work on Avi Video files. Implemented project first loads Avi video file, then initiates shot boundary detection. The algorithm has been tested for the two different videos 1. Shrek_3.Avi (2 minutes of duration) and 2. Tom_jerry.Avi (2 minutes of duration). These video clips are obtained from the internet. Then detected shot boundaries are used as an input to obtain the video transitions effects as abrupt cut and fade in /out. Evaluation can be done by keeping the total number of frames constant from each video. Threshold factor is considered as 1. The number of frames considered is 500. The results have been evaluated for the different block sizes, like 3x3, 5x5, 8x8, 10x10 pixels. In order to measure the performance, the following measures were considered.

1. Precision measure: It is ratio of correct experimental detections over the number of all experimental detections.

Precision Rate = correctly detected / (Correctly detected + Missed detected).

2. Recall measure: It is sensitivity that corresponds to the ratio of correct experimental detections over the number of all true detections.

Recall Rate = correctly detected / (Correctly detected + Errorly detected).

In order to compare the overall performance, F is a measure which combines recall and precision results with equal weight $F(\text{recall}, \text{precision}) = (2 * \text{recall} * \text{precision}) / (\text{recall} + \text{precision})$. The results of the evaluation for the video Shrek_3.avi and Tom_jerry.avi are shown in the Table 1 and 2 as below:

		Histogram based SBD			Euclidean based SBD		
Block size	Transition	PR	RR	F	PR	RR	F
3X3	Cut	67.5	96.42	79.41	92.86	100	96.29
5X5	Cut	68.42	96.29	80.00	92.86	100	96.29
8X8	Cut	79.16	95	86.35	92.86	100	96.29
10X10	Cut	92.86	86.66	89.65	92.86	100	96.29
3X3	Fade in/out	79.41	100	88.52	90	100	94.73
5X5	Fade in/out	77.78	100	87.50	90	100	94.73
8X8	Fade in/out	79.41	96.42	87.09	90	100	94.73
10X10	Fade in/out	95.45	100	97.67	90	100	94.73

Table 1: Shrek_3.avi analysis results

As per the Table 1 results in Shrek_3.avi video, the PR and RR values in histogram based SBD is better as the block size increases. But the PR and RR value in the euclidean based SBD remains same irrespective to the block size. So, the overall performance of the histogram based SBD increases as the block size increases. In case of euclidean based SBD, the overall performance remains the same irrespective of the block size.

		Histogram based SBD			Euclidean based SBD		
Block size	Transition	PR	RR	F	PR	RR	F
3X3	Cut	62.16	85.18	71.87	66.66	100	80
5X5	Cut	63.63	87.50	73.67	66.66	100	80
8X8	Cut	66.66	85.71	74.99	66.66	100	80
10X10	Cut	69.56	88.89	78.04	66.66	100	80
3X3	Fade in/out	71.42	100	83.32	84.61	100	91.67
5X5	Fade in/out	70.45	93.93	80.51	84.61	100	91.67
8X8	Fade in/out	73.68	96.55	83.57	84.61	100	91.67
10*10	Fade in/out	75	100	85.71	84.61	100	91.67

Table 2: Tom_jerry.avi analysis results

As per the Table 1 results in Tom_jerry.avi video, the PR and RR values in histogram based SBD is better as the block size increases. But the PR and RR value in the euclidean based SBD remains same irrespective to the block size. So, the overall performance of the histogram based SBD increases as the block size increases. In case of euclidean based SBD, the overall performance remains the same irrespective of the block size.

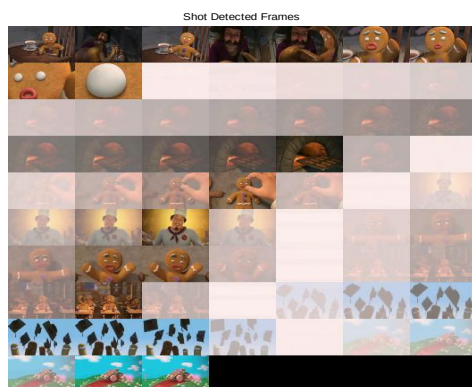


Figure 1: Shot boundaries detected using Histogram based SBD

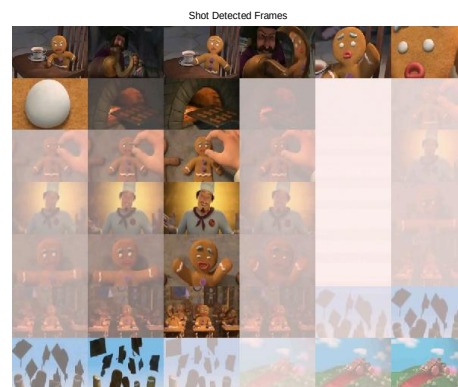


Figure 2: Shot boundaries detected using Euclidean distance based SBD

The detected shot boundaries for the video Shrek_3.avi using the histogram based SBD is shown in the fig 1. The detected shot boundaries for the video Shrek_3.avi using the euclidean based SBD is shown in the fig 2.

IV. CONCLUSION & FUTUREWORK

In this paper, two methods of the shot boundary detection algorithms are analyzed and verified. First the video is segmented into frame, and then employed into histogram difference or the euclidean difference to compute the threshold value. By using this threshold value, shot boundaries are detected. Based on these boundaries, the video transitions have been obtained such as abrupt cut and fade in/out. Experimental results show that, in histogram based method of SBD, the performance increases with respect to the increase in the block size. But in case of euclidean distance based, the performance remains the same irrespective of block size. Overall, the euclidean distance method of SBD gives a better performance against the histogram method of SBD. This summarization tool can be further extended to for feature extraction for the video retrieval and indexing mechanism. The main future enhancement to this paper is the end-end single package development from the video summarization to video indexing and retrieval mechanism.

References

- [1] Sandip T. Dhagdi, P. R. Deshmukh, "Key Frame based Video Summarization Using Automatic Threshold and Edge Matching Rate". International Journal of Scientific and Research Publications, Volume 2, Issue 7, July 2012.
- [2] Rainer Leinhardt, "Comparison of Automatic Shot Boundary Detection Algorithms". Microcomputer Research Labs, Intel Corporation.
- [3] Wael Abd-Elmageed, "Online, Simultaneous Shot Boundary detection and Key frame extraction for Sports videos using Rank Tracing". Institute for Advanced Computer Studies, University of Maryland, ICIP 2008.
- [4] Jun Li, Youdong Ding and Wei Li, Yunyu Shi, "DWT- based Shot Boundary Detection using Support Vector Machine". International Journal of Computer Information Systems and Industrial Management Applications (IJCISIM), Vol 1, pp.214-221, 2009.
- [5] Chun F. Hsu, Mong-Kai Ku, "SVM Based Shot Boundary Detection Using Field Programmable Gate Arrays". Lab of System-On-A-Chip, Dept of Computer Engineering, National Taiwan University
- [6] Weiming Hu, Naianhua Xie, Li Li, Xianglin Zeng and Stephan Maybank, "A Survey on Visual Content based Video Indexing and Retrieval", Volume 41, No.6, November 2011
- [7] Lenka Krulíková, Jaroslav Polec, "An Efficient Method of Shot Cut Detection". World Academy of Science, Engineering and Technology 63 2012.
- [8] Z. Cernekova, C. Nikou, I. Pitas, "Shot detection in Video Sequences using Entropy based metrics". Department of Informatics, University of Thessaloniki, GREECE, IEEE ICIP 2002.