ISSN NO. 2320-5407

Journal homepage: http://www.journalijar.com

INTERNATIONAL JOURNAL OF ADVANCED RESEARCH

RESEARCH ARTICLE

Video Shot Detection using Image Property

Archana V[1], and Dr. Jharna Majumdar[2]

1. MTech, 4th Semester, Department of CSE, Nitte Meenakshi Institute of Technology, Bangalore, India

2. Dean R&D, Professor and Head CSE (PG), Nitte Meenakshi Institute of Technology, Bangalore, India.

Manuscript Info

Manuscript History:

Received: 18 June 2015
Final Accepted: 22 July 2015
Published Online: August 2015

Key words:

Intensity Difference, Image Property, Activity Tracking

****Corresponding Author***

Archana V

Abstract

In the current world, the multimedia information is gaining more importance. It is more difficult to extract the video from huge databases, because the overall structure of the video is complicated. Hence, it is required for efficient and accurate video extracting methods. The objective of this work introduces the various methods of shot detection for various video. Initially the work demonstrates the Global and Local Intensity Difference for Video Shot Detection. The next work aims to demonstrate the Image Properties for video shot detection. Here, the work focuses on to determine which property gives the maximum block-wise difference value, and also the maximum local difference value when a transition occurs and the property giving the high value are used on videos for detecting shots. The next section involves performing the activity detection of walking and jumping sequence in a given video.

INTRODUCTION

The expanded accessibility and the increase in the usage of video have made a necessary requirement for automatic video analysis techniques. Various analysis methods on this content retrieval involve in finding the shot-boundaries in a video automatically. Temporal video segmentation is the beginning step towards an automatic representation of the video sequences. The aim is to partition the video into a smaller units or segment called as shots.

The first step in the analysis and content-based retrieval is to divide a given video into shots. A shot in a video can be outlined as a series of consecutive frames which is taken continuously from a single camera. Hence, the shots are referred to be the basic units for a video indexing, analysis, and classification.

Video shot boundary detection could be a basic step for video analysis based on the content. Video shot detection is the primary and important phase for video indexing, browsing and also in the retrieval applications. This process of shot detection should let the computer to discover the edited positions in the video and should produce the original shot sequences. Therefore the detected shots become the query units in the video retrieval systems.

Shot detection in the video is an initial step in video indexing and retrieval systems. The objectives include dividing the video sequence into manageable shots, identify and classify the various types of transitions in the video.

A figure 1 below depicts that the video sequence is a set scenes and the scenes collection of shots. The shot is a sequence of frames. The boundary between the shots is called as shot transitions. The frames are the smallest units of a video sequence.

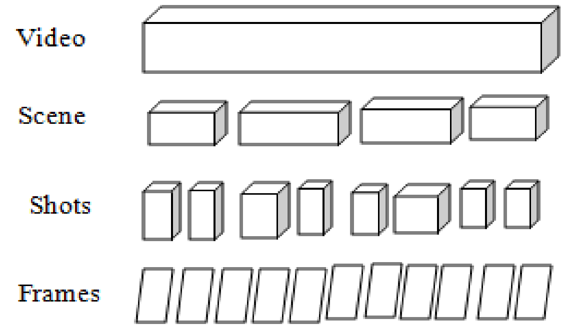

Figure 1: Structure of a video sequence

The basic method involved in the detecting the shots from the video sequence are by calculating the dissimilarities or the discontinuities in the visual contents with respective to the time.

There are of two types of shot transitions namely: hard cut transition and gradual transition. The gradual transition includes various types of transitions such as wipe transition, fade transition, and dissolve transitions. While the abrupt or hard cut transition occurs over a single frame, the gradual transitions happen over multiple frames.

A hard cut is a sudden change in the content of the frame that happens over a single frame. In this case, the first image is a member of the disappearing shot and the second image belongs to next appearing shot.

Figure 2: Hard cut Transition

A fade transition takes place slowly over few frames where there is a change in the brightness which results in black frame or starts with a black frame. During the fade transition, the brightness of the previous shot reaches to zero which is also known as fade-out, and the next shot brightness increases from zero to an image which is known as fade-in. In this type of transition, the previous shot slowly turns out to a blank frame, and the next shot fades in from the blank frame.

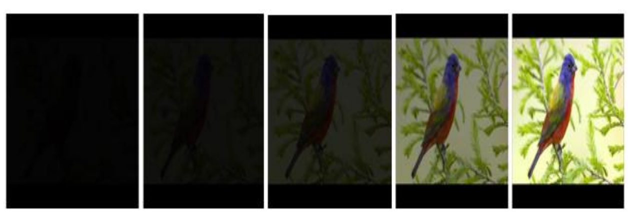

Figure 4: Fade-in transition

Figure 3: Fade-out transition

Dissolve transition occurs in the video when the frame of the first shot slowly disappears and the frames of the second shot brightness increases. The frames involved in the transitions are superimposed on one another.

Figure 5: Dissolve transition

A wipe transition is seen in the video when the second shot frames gets replaced with the pixels of the previous shot frames in a particular pattern, for e.g., a line moves from left to right, or from right to left side of the frames.

Figure 6: Wipe transition

Proposed Method
Intensity Difference
Global Intensity Difference

This is the simple method for detecting shot boundaries. This method finds out pair-wise intensity difference between consecutive frames. The change between the two frames can be determined by comparing the feature difference values of the corresponding intensities in the two frames.

It can be formulated as

$$D(i,i+1) = \frac{\sum_{x=1}^{M}\sum_{y=1}^{N} |F_i(x,y) - F_{i+1}(x,y)|}{M\ X\ N}$$

Where, i and i+1 are two successive frames of height M and width N at the coordinates (x, y) in the frames.

Local Intensity Difference

The block-wise methods make use of the local features of the frames in order to increase the efficiency for the camera and object movement and increase the robustness. The frame is first divided into b number of blocks, and every block value is compared with its corresponding block between the consecutive frames. The difference between two frames **i** and **i+l** is measured by

$$D(i,i+1) = \sum_{k=1}^{b} |DP(i,i+1,k)|$$

Where, the frames are divided into b blocks and DP(i,i+1,k) is the difference of kth block between ith and (i+1) th frame.

Image Properties

Contrast: It is the difference in color that makes an image distinguishable from other images.

$$Contrast = \frac{1}{mn}\sqrt{\sum_{i=1}^{m}\sum_{j=1}^{n}[f(i,j) - \mu]^2}$$

Entropy: The amount of information obtained from an image is measured.

$$Entropy = -\sum_{i=0}^{255} p(i)\log p(i)$$

Spatial Frequency: It measures the overall activity level in an image

$$Spatial\ Frequency = \sqrt{Row\ Frequency^2 + Column\ Frequency^2}$$

$$Row\ Frequency = \sqrt{\frac{1}{mn}\sum_{m=1}^{M}\sum_{n=2}^{N}(F(m,n) - F(m,n-1))^2}$$

$$Column\ Frequency = \sqrt{\frac{1}{mn}\sum_{n=1}^{N}\sum_{m=2}^{M}(F(m,n) - F(m-1,n))^2}$$

Visibility: It measures the amount of clarity that is visible in the image.

$$Visibility = \sum_{m=1}^{M}\sum_{n=1}^{N}\frac{|F(m,n) - \mu|}{\mu^{\alpha+1}}$$

Where μ is the mean value of the Image and α is a constant.

Global Image Properties

The method was initially tested on a set of images to find out the property giving the maximum difference value. The image property visibility difference showed the maximum difference value among other image properties. Hence, the visibility property is applied for video for the shot detection.

Local Image Property Difference

Here, the frames are divided into fixed number of blocks. The image property method was initially tested on a set of images to find out the property giving the maximum local difference between frames. The image property visibility difference showed the maximum difference compared to the other three image properties. Hence the visibility property for the local difference is applied for video for the shot detection.

ALGORITHM

Algorithm: Global Difference Method

1. Read the frames one by one from the input video.
2. Calculate the global difference value for consecutive frames in the video.
3. Calculate the threshold by the following formula.
   Th = mean + α * standard deviation.
   Where α is the alpha value.
4. If the difference obtained is greater than threshold, then we can conclude that the shots are detected.

Algorithm: Local Difference Method

1. Read the frames one by one from the video.
2. Divide the frames into 3x3 blocks.
3. Calculate the Block-Wise difference value for consecutive frames of the video.
4. Calculate the threshold by the following formula.

Th = mean + α * standard deviation.
Where α is the alpha value.

5. If the difference obtained is found to be greater than threshold, then we can conclude that the shots are detected.

The detailed results of shot detection for intensity difference and image property are shown in the experimental results section.

Activity Tracking

Recognizing human activities from video is important applications of computer vision. This application is very useful for video surveillance, human-computer interface, industry, academia, security agencies.

This system has a unique role which observes the persons or the object's when they are under moving state.

The primary assumption to be made in this detection sequence is to capture the video using a fixed camera and determining the activity of a person in the scene.

The two scenarios of activity are considered: human walking activity and human jumping activity.

Detection of walking activity in the scene

The aim of this section is to detect the activity sequences such as a person walking in a scene. The slow movement of a scene can be seen by initially dividing the frame into a fixed block and the changes between the frames are considered using the visibility property. The graph showing almost the same peak is considered as a walking activity of a human.

Detection of jumping activity in the scene

The aim of this section is to detect the activity sequences such as a person jumping in a scene. The jumping movement in a scene can be done as follows: Initially the frames are divided into a fixed block and the block-wise changes are compared between the frames using visibility property. The graph showing uneven peak heights are considered as a jumping activity of a human.

Experimental Results

The proposed methods has been implemented on a Microsoft Visual Studio platform using VC++ programming language.

The proposed methods have been implemented on 5 videos set for each method. Following results shows the shots detected frames for the video which shows the best results.

Intensity Difference

Figure 7: Shot frames of Cut2 video

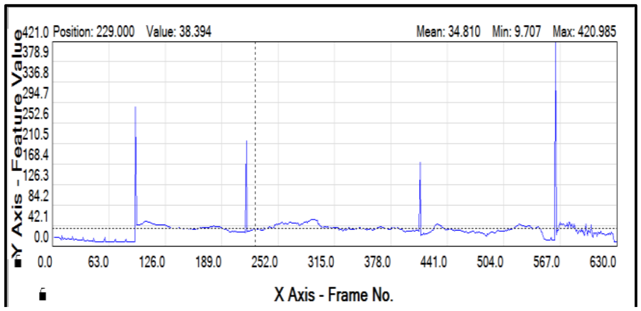

Figure 8: Graph showing the cut shots of Cut2 video

Image Property Difference

Figure 9: Shot frames of wildlife video

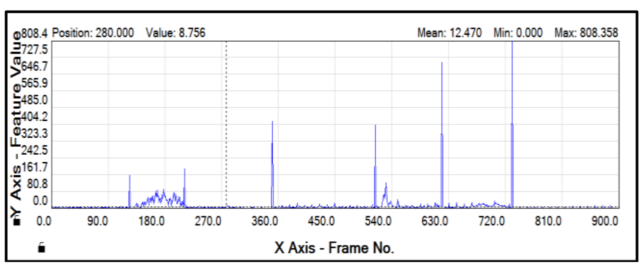

Figure 10: Graph showing the cut shots of Wildlife video

Detecting walking activity in a scene

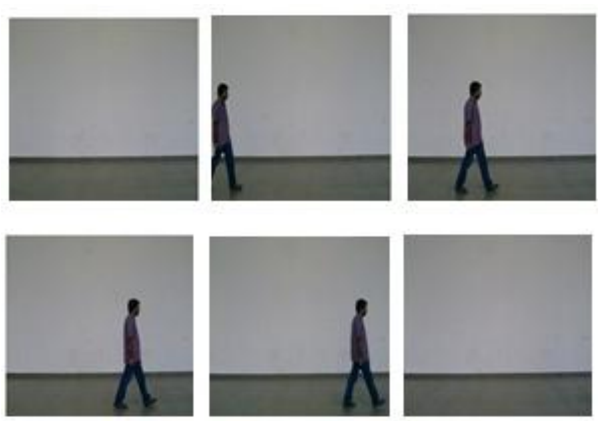

Figure 11: Frame sequence for walking

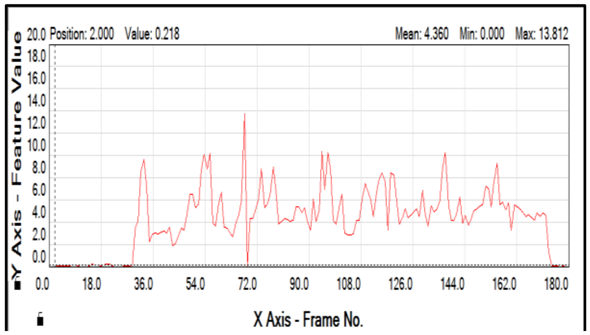

Figure 12: Output showing walking activity of a human

Detecting Jumping activity in a scene

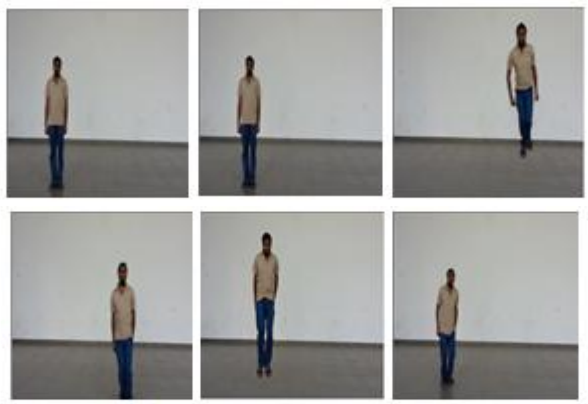

Figure 13: Frame sequence for jumping

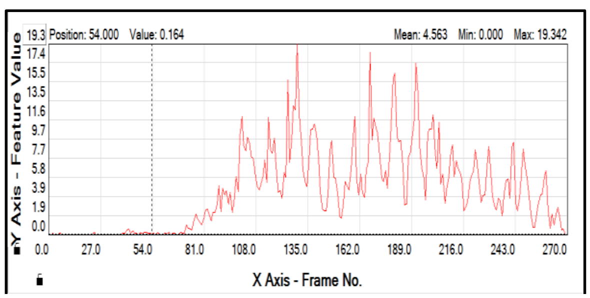

Figure 14: Output showing jumping activity of a human

Performance Analysis

The basic measures for performance evaluation in shot detection are the recall and precision values. Recall is the measure of how much proportion of the correct shots is detected in the video. Precision is the measure of how much of the detected shots are correct. Recall and Precision can be estimated by the following formula:

$$Recall = \frac{No.of\ Correct\ Shots}{No.of\ Correct\ Shots + Missed\ Shots}$$

$$Precision = \frac{No.of\ Correct\ Shots}{No.of\ Correct\ Shots + False}$$

The Table1 shows the recall and precision values for 5 different videos for Intensity Difference

Table 1: Precision and Recall for Intensity Difference

Method	*Video Name*	*Precision (%)*	*Recall (%)*
Global Intensity Difference	Park	100	100
	Cut2	100	100
	Wildlife	86	100
	HurricaneForce	100	100
	Football	88	88
Local Intensity Difference	Park	100	100
	Cut2	100	100
	Wildlife	100	100
	HurricaneForce	100	100
	Football	89	100

Table 2 shows the recall and precision values for 5 different videos for Image Properties.

Table 2: Precision and Recall for Image Property Difference

Method	*Video Name*	*Precision (%)*	*Recall (%)*
Global Visibility Difference	Cut1	80	100
	Cut2	75	75
	Wildlife	60	100
	HurricaneForce	100	63
	Football	78	88
Local Visibility Difference	Cut1	100	100
	Cut2	100	75
	Wildlife	100	100
	HurricaneForce	100	63
	Football	88	88

The above tables show that the local feature is more accurate and efficient than the global feature for maximum videos.

Conclusion

The methods used in this work developed an efficient system to detect the shots in the video sequence. The visibility feature of an Image Property gives the Maximum Difference value when a cut occurs and hence can efficiently detect the hard cuts in the video sequence. The local feature provides more efficiency compared to global feature. The future work includes working on these methods on various types of gradual transitions.

Activity tracking results successfully identifies the person present in the scene and also can determine the type of activity of a person such as walking and jumping activities providing a robust system.

References

Mona Omidyeganeh, Shahrokh Ghaemmaghami, Shervin Shirmohommadi(2011), “Video Keyframe Analysis Using a Segment-Based Statistical Metric in a Visually Sensitive Segment-Based Statistical Metric in a Visually Sensitive,” IEEE Transactions on Image Processing

Yongliang Xiao, Limin Xia(2010), “Key Frame Extraction Based on Connectivity Clustering”, IEEE, Second International Workshop on Education Technology and Computer Science :174 - 177

Pamarthy Chenna Rao and M. Ramesh Patnaik(2014), “Contourlet Transform Based Shot Boundary Detection,” International Journal of Signal Processing, Image Processing and Pattern Recognition, Vol.7, No.4, pp.381-388

Prajesh V. Kathiriya, Dhaval S. Pipalia, Gaurav B. Vasani, Alpesh J. Thesiya, Devendra J. Varanva(2013), " X^2 (Chi-Square) Based Shot Boundary Detection and Key Frame Extraction for Video," International Journal of Engineering And Science, Pp 17-21

Malay Kumar Kundu and Jaydeb Mondal(2012), “A Novel Technique for Automatic Abrupt Shot Transition Detection”, IEEE, Volume 57, issue 1, pp 628-631

B. F. Momin, S. B. Pawar(2014), Key Frame Extraction Using Features Aggregation, International Journal of Recent Development in Engineering and Technology, Volume 2, Issue 1

G G Lakshmi Priya, S. Domnic(2012), “Transition Detection using Hilbert Transform and Texture Features”, American Journal of Signal Processing, pp: 35-40

Ebrahim Peng Taile, Zhang Wenjun(2014), “Robust Shot Boundary Detection from Video Using Dynamic Texture”, IFSA Publishing, pp: 104-109

Dr. Sudeep D. Thepade, Ashvini A. Tonge(2014), “Extraction Of Key Frames from Video Using Discrete Cosine Transform” ICCICCT

Karim M. Mahmoud, Nagia M. Ghane, and Mohamed A. Ismail(2013), “Unsupervised Video Summarization via Dynamic Modeling-based Hierarchical Clustering”, 12th International Conference on Machine Learning and Applications