International Journal of Science and Research (IJSR)
ISSN (Online): 2319-7064
Index Copernicus Value (2013): 6.14 | Impact Factor (2013): 4.438

Video Shot Detection using Texture Feature

Priyanka A R[1], 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

Abstract: - ***The multimedia data has the need for finding out some of the effective and robust methods mainly for storing the information in the database. Also, in its beginning step, accessing the content of the video necessitates the searching of video information. Hence, a huge amount of methods needs to be proposed in order to solve the main issues in the detection of shots. In this paper, first we demonstrate the Global and Local Pixel-wise Difference. Here, the main aim is to prove that the Local Pixel-wise difference method can accurately identify the exact position of change when a shot-transition occurs when it is compared to Global Pixel-wise difference. Secondly, we demonstrate the GLCM for Video-Shot-Detection in order to determine which feature gives maximum global difference and which feature gives maximum local difference when a shot transition occurs. Finally, we demonstrate the detection of the slow movement and jumping activity in a scene.***

Keywords: Video Shot Detection, Global Pixel-Wise Difference, Local Pixel-Wise Difference, GLCM, Activity Tracking.

1. Introduction

In the interactive media environment, the utilization of advanced information is expanding quickly and the tools to handle extensive volumes of video information are needed. In recent years, the use of video based information is increasing a lot. Because the value of the devices that are used for storing are reducing and improvised techniques, the digital video are obtained at a faster rate.

Video information is accessible and is being used in most of the application in various areas such as the business, digital marketing, electronic business, broadcasts, video-on demand recreation, etc.

The digital-video field is very much advanced and also within a couple of years the accessibility of resources has increased rapidly. In most of the cases, the digital-video can be easily accessed from anywhere with an increase in the accessibility. This is kind of video information is not at all helpful for older forms of data access, surfing and searching. Hence, many methods have been used for changing the video knowledge into a lot of compact forms and take only the ones that are very useful information from them. These kinds of operations will provide as a first stride for a number of variety of information access undertakings. At this point, the analysis of the video which are at a high-level are improved by using the basic methods which are developed to help them. The initial tasks are detection of the shot-boundary and representing the video condensed in the condensed form.

Video shot boundary detection is the elementary stride of categorization and surfing applications. The general objective is to mainly segment a video sequence into its constituent shots and then to identify the different transitions among the shots in the video sequence. The initial step in the video processing method is to divide the video into the shots.

A shot in a video is outlined as a series of interconnected successive frames taken continuously by a camera. A scene is outlined as a collection of one or additional adjoining shots that focus on an object or objects of interest.

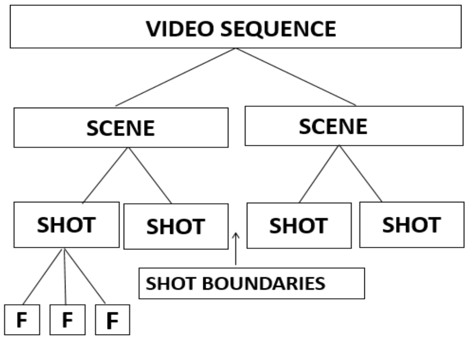

Figure 1: Video divided into scene, scenes into shots and shots into frames.

The above figure depicts how the video sequence is divided into a set of scenes and how the scenes are further divided into shots. Finally, the shots are further divided into frames. The boundary between the shots is called as shot-transitions. The frames are the smallest unit of a video.

The transitions are of two types that can occur between shots, they are abrupt change and gradual transition. The shot boundaries must be analyzed in order to detect whether they have any gradual transitions in the midst of the frames is a very difficult task. The gradual change is extended over a set of frames and is divided into fade, wipe or dissolve transition.

A hard cut is said to have occurred when there is an abrupt change in the scene from the current frame to the next frame.

Figure 2: Abrupt Transitions -Hard-Cut Transition

In Fade, the brightness of the image slowly changes within a scene and it usually results in fading in or fading out of pixel towards complete black or white.

Figure 3: Gradual Transition –Fade out Transition

Figure 4: Gradual Transitions –Fade in Transition

A Wipe is detected when the pixels from the changed frame replace those of the current frame either in Horizontal or Vertical fashion. This traverses from the one side of the frame to another.

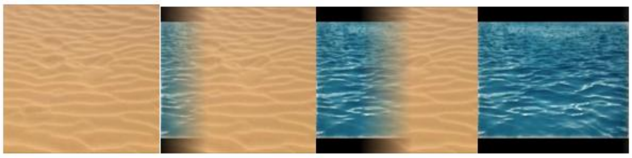

Figure 5: Gradual Transitions –Wipe Transition

A Dissolve is detected when there is a blending of two different images over a period of frames. Here, the current image gradually fades while the next image replaces the previous image.

Figure 6: Gradual Transitions –Dissolve Transition

2. Proposed Method

Algorithm should be designed with the reduction of number of false detections in the presence of object and free from the complexity of computational time in mind. Here, we propose 3 methodologies for detecting the hard cut transitions using the concept of global and local features.

2.1 Global Pixel-wise Difference

This is one of the techniques for detecting the shot boundaries. This technique calculates the pair-wise difference between pixel-wise intensities among consecutive frames. The change between the frames is identified by comparing the difference values of intensity of the corresponding pixels in the consecutive frames.
It can be calculated as follows

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

Where i and i+1 are the two consecutive frames with dimension MXN, $P_i(x,y)$ is the intensity value of the pixel at the coordinates (x,y) in frame i and $P_{i+1}(x,y)$ is the intensity value of the pixel at the coordinates (x,y) in frame i+1.

2.2 Local Pixel-wise Difference

Alternative to the Global Pixel-wise difference, the block-based method uses the local characteristics mainly to increase the efficiency for camera and object movement. Here, all the frames are first divided into 'b' no of blocks, and each block of the current frame is compared with its corresponding block in the next frame. The difference between two frames **i** and **i+l** is computed by the following formula

$$D(i,i+1) = \sum_{k=1}^{b} |DP(i,i+1,k)| \qquad (2)$$

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

2.3 Grey Level Co-occurrence Matrix

The GLCM is a square matrix that specifies some of the characteristics like distribution of the gray-values in the texture image. It shows how many times intensity value also called as the reference point with a value k occurring in a specific direction with a intensity value called as the neighbor point with the value l. Every element in the matrix is the number of times a pair of intensity having the value k and a intensity having the value l are at a distance d in relation with each other. The relationship between two neighboring pixels can be defined in different ways in different angles. The default angle is between a pixel and its neighbor to its right. Finally, it can be said that for an image F of size KxK, the elements of a gray matrix M_{CO} for an angle d= (dx, dy). It is as shown below

$$M_{CO} = \sum_{x=1}^{k}\sum_{y=1}^{k} \begin{cases} 1, if\ I(x,y) = i\ and\ I(x+dx, y+dy) = j \\ 0, otherwise \end{cases} \qquad (3)$$

2.3.1 Texture Features of GLCM

Energy (Angular Secondary Moments): It is also called as Uniformity or Energy and it is a sum of squares of entries in the matrix.

$$\text{Energy} = \sum_{i,j=0}^{N-1} P_{i,j}^{2} \qquad (4)$$

Entropy: Entropy specifies the amount of information of an image which is required for the image compression.

$$\text{Entropy} = \sum_{i,j=0}^{N-1} P_{i,j}(-\ln P_{i,j}) \qquad (5)$$

Contrast: Contrast specifies local grey-level changes in the grey-level co-occurrence matrix.

$$\text{Contrast} = \sum_{i,j=0}^{N-1} P_{i,j}(i-j)^2 \qquad (6)$$

Maximum Probability: It records the center pixel inside the window and the highest pixel Pij value that is found within the window.

$$\text{Maximum Probability} = Max\,(P_{i,j}) \qquad (7)$$

Homogeneity: It specifies the values by taking the inverse of the Contrast value weight, and the values are decrease exponentially far from the diagonal.

$$\text{Homogeneity} = \sum_{i,j=0}^{N-1} \frac{P_{i,j}}{1+(i-j)^2} \quad (8)$$

Variance: It is a measure of dispersion of the values around the mean value and it is similar to the entropy feature.

$$\text{Variance} = \sigma_i^2 = \sum_{i,j=0}^{N-1} P_{i,j}(i-\mu_i)^2 \qquad \sigma_j^2 = \sum_{i,j=0}^{N-1} P_{i,j}(j-\mu_j)^2 \quad (9)$$

$$\text{Where} \quad \mu_i = \sum_{i,j=0}^{N-1} i(P_{i,j}) \qquad \mu_j = \sum_{i,j=0}^{N-1} j(P_{i,j})$$

2.3.2 Global GLCM Difference

This concept was initially tried on images for various features to find out the feature giving the maximum difference when a shot transition occurs and that feature was run on video for Shot Boundary detection.

2.3.3 Local GLCM Difference

This method was first tested on images for various texture features to find out the feature giving the maximum block-wise difference when a shot transition occurs and that feature was run on video for Shot Boundary detection.

2.4 Algorithms

2.4.1Shot Detection using Global Feature

1. Extract the frames from the video.
2. Calculate the global difference between the currently read frames with the previously stored frame.
3. Calculate the Mean and Standard Deviation of difference values and calculate the threshold (Th) using the mean and standard deviation as shown below
   $$\text{Th} = \text{Mean} + \alpha * \text{Standard deviation} \quad (10)$$
4. When the difference that is calculated above is found to be higher than the threshold, then the shots are detected.

2.4.2 Shot Detection using Local Feature

1. Extract the frames from the video.
2. Divide each of the frames into 3x3 blocks.
3. Calculate the local difference between the currently read frames with the previously stored frame.
4. Calculate the Mean and Standard Deviation of difference values and calculate the threshold as shown below.
   $$\text{Th} = \text{Mean} + \alpha * \text{Standard deviation} \quad (11)$$
5. When the difference that is calculated is found to be higher than the threshold, then the shots are detected.

The above algorithms are implemented for both Pixel Difference and GLCM for variance feature and its results are shown in the experimental results section.

2.5 Activity Tracking

Tracking people in known environments has recently become an active area of research. Robust, multi-person tracking systems have possible use in a wide range of applications, including smart videoconferencing systems, surveillance for security and/or site evaluation, as well as providing location and context features for human-computer interaction. The main objective of this system is to detect the persons or objects that are under motion.

2.5.1 Detection of slowly moving part of the scene

There are some of the considerations in this method: The image sequence must be recorded with a fixed camera and using this sequence the person's walking activity can be detected. Here, the frames I_1 to I_n are analyzed. The detection is done mainly on the basis of the peak obtained in the graph after taking the local difference. The experimental results shows the peak obtained for the walking activity remains almost constant.

2.5.2 Detection of jumping activity in the scene

The most important considerations in this method are as follows: The sequences of the image must be recorded with a fixed camera and using this sequence the person's jumping activity can be detected. Here, initially the frames F_1 to F_n must be analyzed. The detection is done mainly on the observation of the peak obtained in the graph after taking the local difference. The experimental results shows that the peak obtained for the jumping activity varies as the person is jumping.

3. Experimental Results

The proposed method has been implemented on 5 videos namely cut1, cut2, wildlife, park, football and entertainment. Following shows the observations of the shot detected frames for the above mentioned methods.

Figure 7: Shot detected using Pixel difference

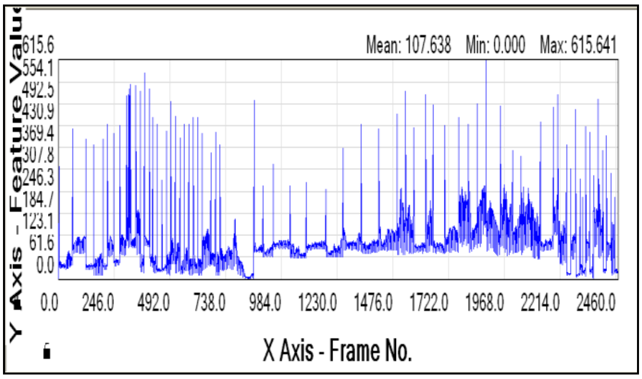

Figure 8: Shows the graph of pixel difference values where the peaks indicate the shot change

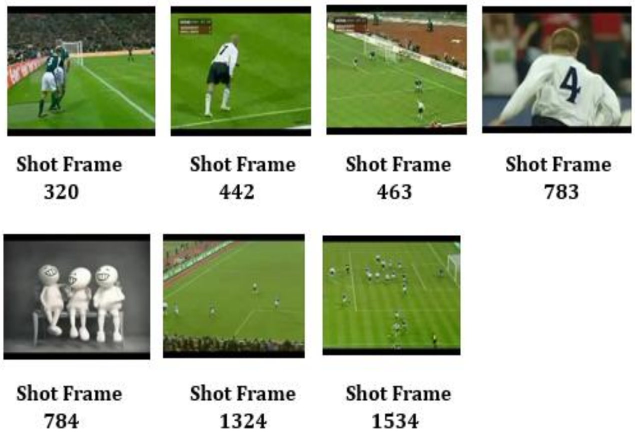

Figure 9: Shot detected frames for variance feature of GLCM

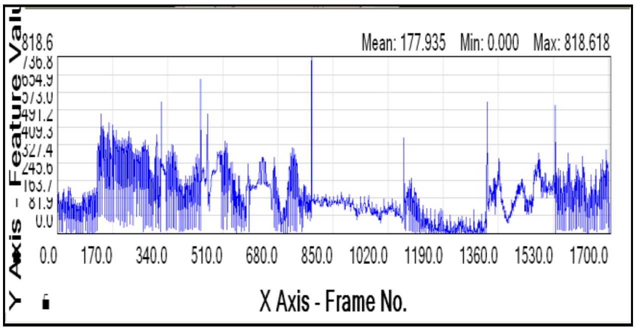

Figure 10: Shows the graph of GLCM for variance feature values where the peak indicates the shot change

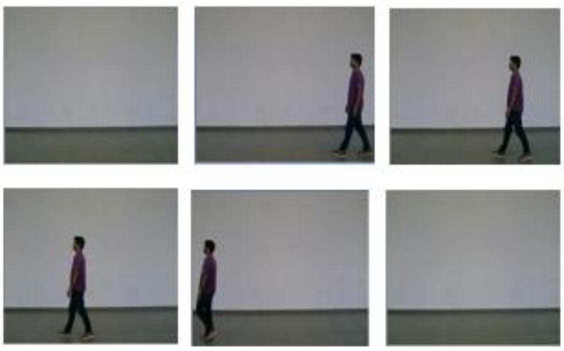

Figure 11: Sample test images of walking activity

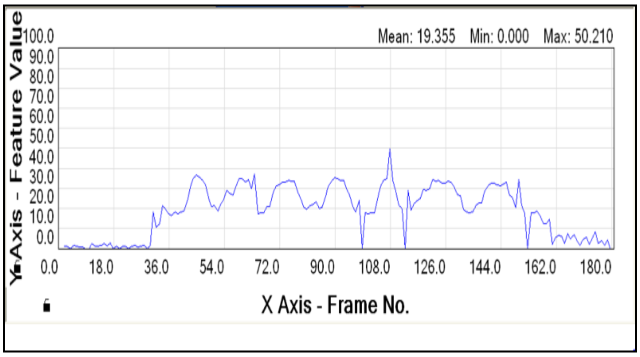

Figure 12: Graph showing the walking activity

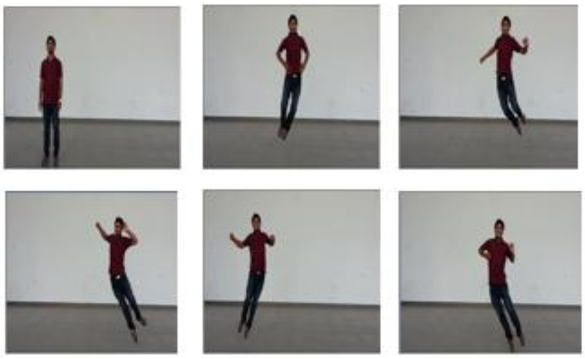

Figure 13: Sample test images of jumping activity

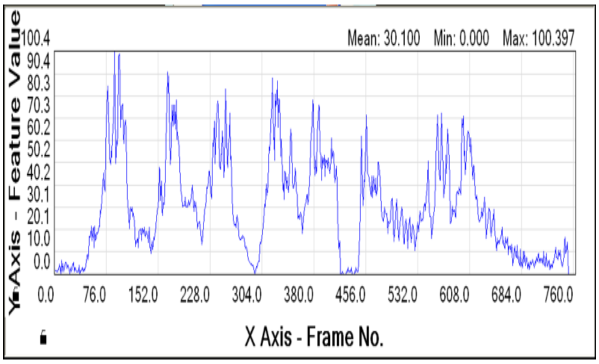

Figure 14: Graph showing the jumping activity

A. Implementation Details

The proposed method has been implemented on a Microsoft Visual Studio platform using VC++ programming language. Windows XP and Windows 7 operating system have been used.

B. Performance Analysis

To evaluate the proposed method precision and recall were used. The recall represents the detection ability of the proposed algorithm and the precision represents its accuracy. The equations for recall and precision are

$$Recall = correct/(correct + Missed) \quad (12)$$

$$Precision = correct/(correct + false\ positve) \quad (13)$$

Table 1: Precision and Recall for Pixel Difference Method

Name of the Video	*Global Pixel Difference*		*Local Pixel Difference*	
	Recall (%)	*Precision (%)*	*Recall (%)*	*Precision (%)*
Entertainment	92	100	98	100
Wildlife	100	86	100	100
Cut1	100	100	100	100
Football	88	88	100	86
Cut 2	100	100	100	100

From table 1, it is seen that local pixel difference method gives better results when compared to global difference methods when a shot transition occurs.

Table 2: Precision and Recall for GLCM Method

Name of the Video	Global GLCM difference		Local GLCM difference	
	Recall (%)	Precision (%)	Recall (%)	Precision (%)
Wildlife	67	80	67	80
Cut2	100	100	100	100
Cut 1	100	75	100	75
Football	100	75	88	100
Entertainment	87	97	95	100

From the above analysis results, it is seen that local GLCM difference method gives better results when compared to global GLCM difference methods when a shot transition occurs.

4. Conclusion

A unique approach for detecting the hard cut transitions has been introduced in this paper. This methodology uses local and the global features to detect the hard cuts. This introduces an efficient and robust system for detecting video scene changes which is an essential task in fully content analysis systems. Experimental result shows that the shots are accurately detected. Future work involves the detection of various types of transitions such as fade, wipe and dissolve and detection of various types of activities.

References

[1] Arun Gautam "Key Frame Selection from video based on weighted Minkowski Distance", Proceedings of 5th SARC-IRF International Conference, 2014.

[2] Sowmya R, Dr.Rajashree Shettar, "Analysis and Verification of Video Summarization using Shot Boundary Detection ", American International Journal of Research in Science, Technology, Engineering & Mathematics, 2013.

[3] Mei huang, Ling Xia, Jin Zhang, Hui Dong," An integrated scheme for Video key frame Extraction", 2nd International Symposium on Computer, Communication, Control and Automation,2013.

[4] Ebrahim Asadi, Nasrolla Moghadam Charkari, "Video Summarization Using Fuzzy C-Means Clustering", 20th Iranian Conference on Electrical Engineering (ICEE), 2012.

[5] K. Koumousis, V. Fotopoulos, A. N. Skodras," A New approach to Gradual Video Transition Detection", IEEE, 2014.

[6] Kai Jin, Hong-cai Feng, Qi Feng, Chi Zhang," Shot Boundary Detection Algorithm Based on Multi-Feature Fusion", ICCSEE, 2013

[7] P. Mahesh Kumar," Shot Boundary Detection in Videos using Graph Cut Sets", International Journal of Emerging Trends in Technology Science & Engineering, 2014.

[8] J Quellet, V Randrianarisoa "To Watch or not to Watch: Video Summarization with Explicit Duplicate Elimination", IEEE, Canadian Conference on Computer and Robot Vision, 2011.

[9] K. S. Mohanty, P. Kanungo, "Automatic Cut Detection Based Video Segmentation", IEEE, 2013.

[10] Huayong Liu, Huifen Hao, "Key Frame Extraction Based on Improved Hierarchical Clustering Algorithm", IEEE, 2013.