

International Journal of Innovative Research in Science, Engineering and Technology

(An ISO 3297: 2007 Certified Organization)

Vol. 4, Issue 8, August 2015

Video Shot Detection using Distance Measure

Kritika Gupta¹, Dr. Jharna Majumdar²

M.Tech, 4th Semester, Department of CSE, Nitte Meenakshi Institute of Technology, Bangalore, India¹

Dean R&D, Professor and Head CSE (PG), Nitte Meenakshi Institute of Technology, Bangalore, India²

ABSTRACT: Videos are becoming popular in today's world. Multimedia videos are capturing in large numbers from various areas, here the most difficult task is to extract the video from large database based on the requirement or needs, because the overall structure of video is complicated. Hence an efficient and accurate video extracting method is required. In this paper, we introduce the methods for video shot detection using Distances Measures in two steps. First, we calculated the distance between two consecutive frames. Afterwards, the distance is compared with threshold for detecting the shots. Comparison between the various distance measures for video shot detection is done in order to determine which distance measure gives the accurate result when a shot transitions occurs. Finally we demonstrate about the detection of slow and fast movement like walking and jumping activity in the scene by the use of Best Distance Measure declared.

KEYWORDS: Distance Measure, Local /Global Pixel Difference, Activity Tracking, Video Shot Detection.

I. INTRODUCTION

In the current world, the multimedia information is gaining its importance day by day. Multimedia videos are capturing in various areas. Traditionally, videos were created only by a limited number of people say producers. But now, the common people as well can afford the video capturing devices and using it with simplicity, as a result of which there is an increased amount of user generated videos. With the advent of digital transmission, plenty of digital content like movies, news video, television shows and sports is widely available. As result of that searching for the content of interest from such a large collection has become a tedious task. For effective use of this, user wants to have a proper control over the video. To resolve the complexity of the problem as a result, many video summarization and indexing techniques are being implemented. This kind of Application needs the videos to be spitted into smallest manageable elementary unit called shots;

The video shot boundary detection becomes the most essential task in such kind of applications. Once the video shots are available using described method from the overall video, the keyframes from each shot can be extracted [1-2]. In [3] author describes the method for video shot detection using histogram differences and local image descriptor. In [4] various similarity and dissimilarity measure has been described for finding the difference between consecutive image. Bhattacharya distance measure is used for finding the similarity between images and author describes here how Bhattacharya give accurate result as compare to other distance measures [5]. In [6-7] various methods of tracking an object and for identifying the activity of object are implemented.

Video Shot boundary detection is just like an opening step for video indexing and retrieving the required data. It is mainly depends on the visual discontinuities occurs between the frame called the transitions, this discontinuities can be found throughout scene. These discontinuities can also be called as transitions shown in fig 1; which are categorized into two types: abrupt: Hard cut(a) and gradual: Fade Out(b), Fade In(c), Wipe(d), Dissolve(e).

A **Hard cut** is whenever the sudden change occurs between the frames of video that we can declare cut occurs.

International Journal of Innovative Research in Science, Engineering and Technology

(An ISO 3297: 2007 Certified Organization)

Vol. 4, Issue 8, August 2015



In gradual transitions the two shots area unit are combined using some chromatic effects that can be step by step replacement of one shot by another.

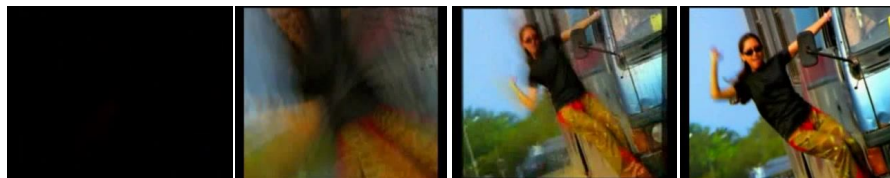
Gradual transitions are of different types they are-

A **Fade Out** is occurs when frame go on fading out until the original content of the shot fully vanishes. What remains may be altogether black frame.



(b)

A **Fade In** occurs when the frame appear into position from a completely dark sequence. The each successive frame becomes brighter in each step.



(c)

A **Wipe** occurs when the frames of the primary shot during this transition are replaced by frames in the second shot in the form of a line.



(d)

A **Dissolve** is a mix of a fade-out transition and fade-in transition wherever the dark sequence is replaced by another shot.



International Journal of Innovative Research in Science, Engineering and Technology

(An ISO 3297: 2007 Certified Organization)

Vol. 4, Issue 8, August 2015



(e)

Fig 1:Types of Transitions: (a) Abrupt: Hard cut,(b) Gradual: Fade Out,(c)Fade In,(d)Wipe,(e)Dissolve

II. DISTANCE MEASURE

Similarity or Dissimilarity measures are at the heart of many image registration methods. Measuring the Similarity or Dissimilarity present between the consecutive images has central importance over low-level computer vision. Similarity or Dissimilarity measures were compared for accuracy under noise and changes in image geometry and intensity. In order to identify the separation between various frames based on the particular selected features, an appropriate and accurate Distance Measure need to be chosen.

Dissimilarity Measure:

1) Minkowski Form Distance

$$d(H_1, H_2) = \left(\sum |f_i(I) - f_i(J)|^p \right)^{\frac{1}{p}} \quad (1)$$

The value of 'd' lies between 0 and N and for a perfect match, value of 'd' should be close to N.

2) Chi-Square Distance

$$d(H_1, H_2) = \sum \frac{(H_1(I) - H_2(I))^2}{(H_1(I))} \quad (2)$$

The value of 'd' lies between 0 and N and for a perfect match, value of 'd' should be close to N.

3) Kullback-Leibler Distance

$$d(H_1, H_2) = \sum_i H_1 \log \frac{H_1}{H_2} \quad (3)$$

The value of 'd' lies between 0 and N and for a perfect match, value of 'd' should be close to N.

4) Jeffery Divergence

$$d(H_1, H_2) = \sum_i \left(H_1 \log \frac{H_1}{m_i} + H_2 \log \frac{H_2}{m_i} \right) \quad (4)$$

$$m_i = \frac{H_1 + H_2}{2}$$

The value of 'd' lies between 0 and N and for a perfect match, value of 'd' should be close to N.

5) Manhattan Distance

$$d(H_1, H_2) = \sum H_1(i) - H_2(i) \quad (5)$$

International Journal of Innovative Research in Science, Engineering and Technology

(An ISO 3297: 2007 Certified Organization)

Vol. 4, Issue 8, August 2015

The value of 'd' lies between 0 and N and for a perfect match, value of 'd' should be close to N.

Similarity Measures:

1) Swain and Ballard

$$S(I, j) = \frac{\sum_{i=1}^n \min f((i), f(j))}{f(j)} \quad (6)$$

The value of 'd' lies between 0 and 1 and for a exact match, value of 'd' should be close to 1.

2) Intersection Distance

$$d(H_1, H_2) = 1 - \sum_i \min (H_1(i), H_2(i)) \quad (7)$$

The value of 'd' lies between 0 and 1 and for a exact match, value of 'd' should be close to 1.

3) Bhattacharya Distance

$$d(H_1, H_2) = \sqrt{1 - \frac{\sum \sqrt{H_1 \cdot H_2}}{\sqrt{\sum H_1(i) \sum H_2(i)}}} \quad (8)$$

The value of 'd' lies between 0 and 1 and for a exact match, value of 'd' should be close to 1.

III. PROPOSED METHOD

3.1 Pixel-Difference

3.1.1 Pixel-Difference using Global Features

This is one of the simplest methodologies for determining shot boundaries. This methodology finds out pair-wise intensities difference between of consecutive frame. The change between the two frames can be observed by comparing the differences value of intensity of corresponding pixels in the consecutive frames.

It can be formulated as

$$D(i, i + 1) = \frac{\sum_{x=1}^M \sum_{y=1}^N |P_i(x, y) - P_{i+1}(x, y)|}{M \times N} \quad (9)$$

where i and i+1 are two successive frames with dimension MxN, $P_i(x, y)$ is, the intensity value of the pixel at the coordinates (x, y) in frame i.

If the difference calculated is greater than the specified threshold, than a shot boundary is assumed.

3.1.2 Pixel-Difference using Local Features

Alternative to the Pixel-Difference based on global feature, block-based approaches uses local characteristic of image mainly to increase the efficiency for camera and object movement. Here every frame is first divided into b no of blocks, and each block of current frame is compared with its corresponding block value in the next followed frame. The difference between two frames i and i+1 is calculated by the following formula given below.

$$D(i, i + 1) = \sum_{k=1}^b |DP(i, i + 1, k)| \quad (10)$$

International Journal of Innovative Research in Science, Engineering and Technology

(An ISO 3297: 2007 Certified Organization)

Vol. 4, Issue 8, August 2015

where every frame is divided into b no of blocks and $DP(i,i+1,k)$ indicated difference of k th block between i th and $(i+1)$ th frame.

If the calculated block wise difference is greater than the threshold, than a shot boundary is assumed.

3.2 Distance Measure

3.2.1 Distance Measure using Global features

This method used various distance measures described in section (II) is calculated as image feature between consecutive frames. This method was initially executed on images for various distance measure to find out which distance measure giving the maximum difference when a shot transition occurs and that distance measure was run on video for Shot Boundary Detection.

Algorithm

Step1: Read the Video containing N frames.

Step2: Apply Minkowski Form Distance between the corresponding pixels of two consecutive frames $N-1$ and N where N is frame.

Step3: Calculate the Absolute sum of the distance value of the consecutive frames.

Step4: If the difference calculated is greater than specified threshold, than the shot boundary is assumed.

3.2.2 Distance Measures using Local feature

This method was first executed on images for various distance measures described in section (II) and finding out which distance measure giving the maximum block-wise difference and at which block. It identifies the maximum change at the part of the frame. When a shot transition occurs and that distance measure was run on video for Shot Boundary Detection.

Algorithm

Step1: Read the Video containing N frames.

Step2: Divide each frame of the video into 3×3 blocks.

Step3: Apply Minkowski form Distance between corresponding block of two consecutive frames $N-1$ and N where N is frame number.

Step4: Identify the block which gives the maximum difference among all the blocks.

Step5: If the calculated block wise difference is greater than the threshold, than a shot boundary is assumed.

3.3 Activity Tracking

Activity tracking is one of the most important tasks in a wide range of applications in computer vision, such as video analytics, human computer interaction, and medical.

Given the position and size of a target object in a frame of a video, here the goal is tracking the states of the target in the subsequent frames.

An activity tracking system has a well-defined role and this is mainly to observe the persons or objects when these are under moving.

3.3.1 Detection of slow movement in the scene

The aim of this section is to detect the activity of a person walking in a scene using the methods described in the section 3.2.2. The slow movement of a scene can be seen by initially dividing the frame into a fixed block-wise and comparing the block-wise changes in the frame. The peak of the graph in a particular block indicates that the person has moved in the corresponding blocks.

3.3.2 Detection of fast activity in the scene

The aim of this section is to detect the activity of a person jumping in a scene using the methods described in the section 3.2.2. 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. The peak in a particular block indicates that the person has jumped in the corresponding blocks.

International Journal of Innovative Research in Science, Engineering and Technology

(An ISO 3297: 2007 Certified Organization)

Vol. 4, Issue 8, August 2015

IV. EXPERIMENTAL RESULTS

In this Section, Two different video clips, “Hurricane.avi” and “Entertainment.avi” are used as input to proposed system. In our proposed system audio parts of the video are neglected. Only visual content changes in the videos are mainly concentrated to detect the shot frames. The algorithm is implemented in VC++ 6.0

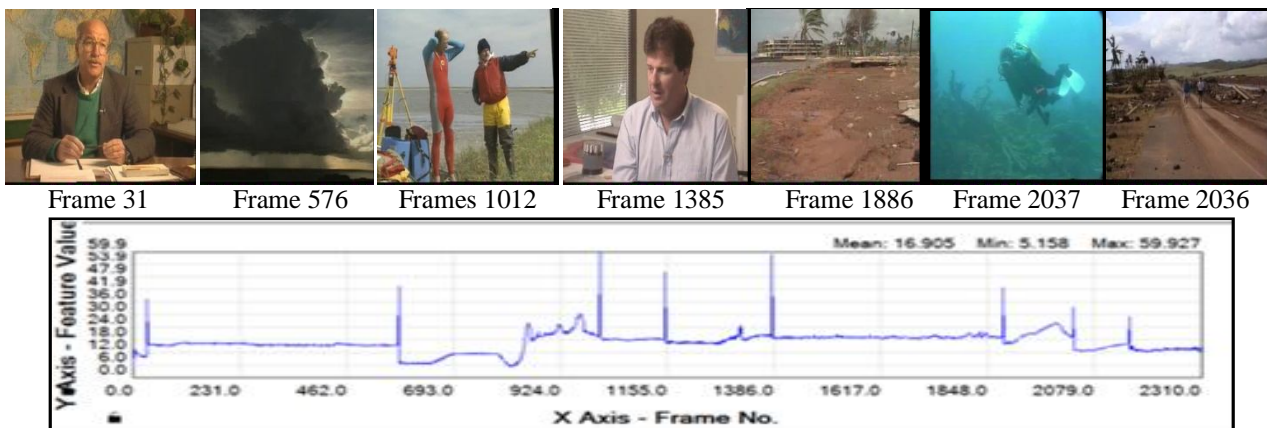


Figure 6: Shot Detected frame and graph using Pixel Difference using Local Feature

The above figure shows shots detected from the Hurricane.avi using pixel Difference (Local feature). There are 7 shots present in the video.

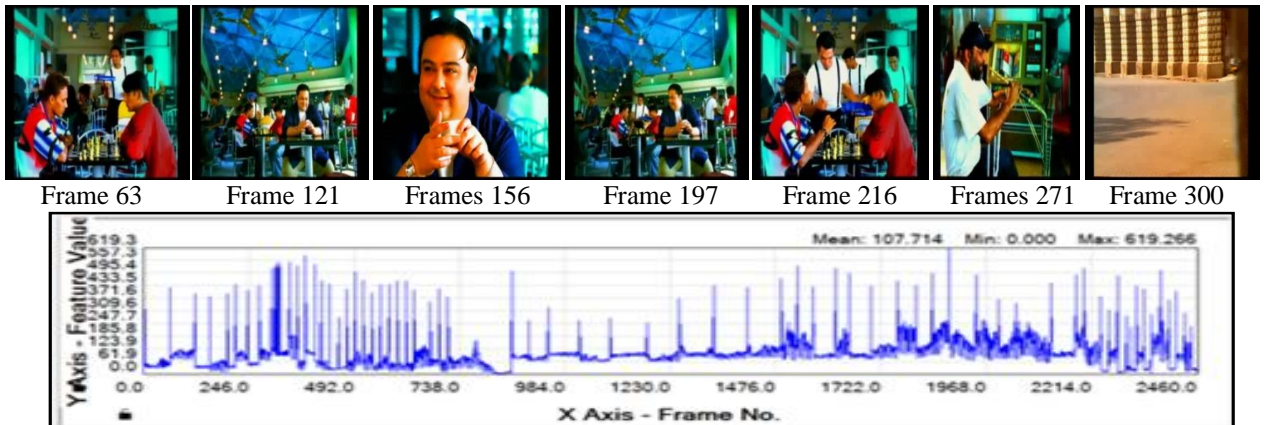


Figure 7: Shot Detected frame and graph using Minkowski form Distance using Local Feature

The above figure shows the shots detected from the Entertainment using Minkowski form Distance (Local feature). There is 62 shots present in the video 7 sample frame are shown above.



International Journal of Innovative Research in Science, Engineering and Technology

(An ISO 3297: 2007 Certified Organization)

Vol. 4, Issue 8, August 2015

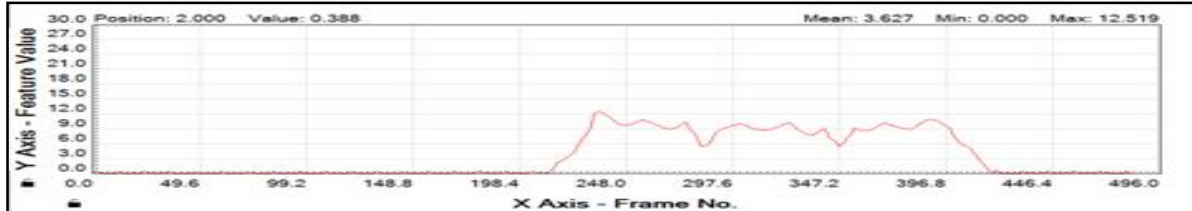


Figure 8: Activity graph of Walking Movement using Minkowski form Distance

The above figure are sample input frame used for proposed method and shows walking activity graph of human, plotted using Minkowski form distance (Local feature).

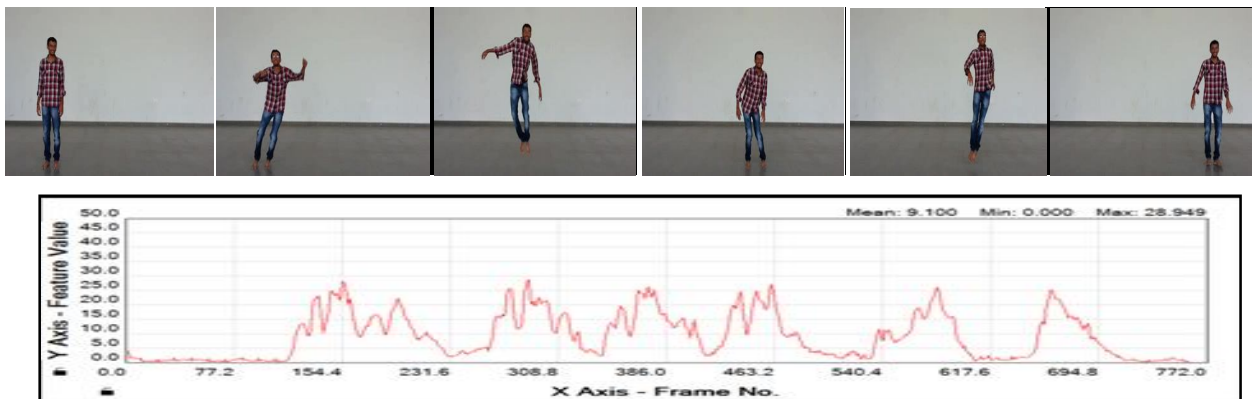


Figure 9: Activity graph of Jumping Movement using Minkowski form Distance

The above figure are sample input frame used for proposed method and shows jumping activity graph of human, plotted using Minkowski form distance (Local feature).

A. Performance Analysis

Our proposed method had been tested over five Different videos which are detailed in section 3.1 and 3.2. To measure this performance, accuracy has to be calculated. Accuracy is measured with the help of Precision and Recall which is done for five different types of video.

Recall(R): (Correct)/(Correct + Miss)

Precision (P): (Correct)/(Correct + False)

Video	Minkowski form Distance (Global)		Minkowski form Distance (Local)	
	Recall	Precision	Recall	Precision
Cut1 (%)	100	100	100	100
Cut2 (%)	100	100	100	100
Wildlife (%)	100	100	100	85
Hurricane (%)	87	100	87	100
Park (%)	100	100	100	100

Table1: Precision and Recall (Quality parameter) Calculation for Minkowski form Distance using local feature.

Video	Pixel Diff (Global)		Pixel Diff.(Local)	
	Recall	Precision	Recall	Precision
Cut1 (%)	100	100	100	100
Cut2 (%)	100	100	100	100
Wildlife (%)	100	100	100	100
Hurricane (%)	100	75	87	100
Park (%)	100	100	100	100

Table2: Precision and Recall (Quality parameter) Calculation for Pixel Difference using local feature

International Journal of Innovative Research in Science, Engineering and Technology

(An ISO 3297: 2007 Certified Organization)

Vol. 4, Issue 8, August 2015

AS=Actual Shots, SD =Shot Detected, MS=Missed shots,
FS=False Shots, CS=Correct Shots.

The above Table 1 shows the analysis on Minkowski form Distance using Global and Local feature with 5 videos and Table 2 shows the analysis on Pixel Difference using Global and Local feature in both the cases we can observe the quality parameter value is high in case of local feature, indicates the Local feature of image give more efficient result than global feature.

V. CONCLUSION

This work introduces the comparative study about the various methods for video shot change detection. This an efficient and robust method for detecting the video scene changes, and for content analysis systems and tracking the object. It is shown from the above results that pixel-difference using local feature can identify the exact position of change when a shot transition occurs when compared to pixel-difference using global feature. In the same manner The Minkowski form distance gives the Maximum Difference value when a cut occurs. Hence it efficiently detects the hard cut. Activity tracking results are successfully identifies with the declared best distance measure that identifies the exact block where the object is present or objects are moved. In future would like to implement other method of video shot detection and methods for tracking the object.

REFERENCES

- [1] Mr. Saurabh Thakare("Intelligent Processing and Analysis of Image for Shot Boundary Detection"), International Journal of Emerging Technology and Advanced Engineering, Volume 2, Issue 2, pp.2250-2459, 2012.
- [2] Rathod, Dipali. A. Nikam2 ,Ganesh, "An Algorithm for Shot Boundary Detection and Key Frame Extraction Using Histogram Difference", International Journal of Emerging Technology and Advanced Engineering, Volume 3, Issue 8, pp. 2250-2459, 2013.
- [3] Youssef Bendraou, Fedwa Essannouni, Driss Aboutajdine1 and Ahmed Salam 2, "Video shot boundary detection method using histogram differences and local image descriptor" IEEE, 978-1-4799-4648-8, pp.665 – 670, 2014
- [4] Prof. A. Ardeshir Goshtasby, "Similarity and Dissimilarity Measures, Image Registration, Advances in Computer Vision and Pattern Recognition," chapter 2, 2012.
- [5] N. A. Thacker, F. J. Aherne and P. I. Rockett, "The Bhattacharyya Metric as an Absolute Similarity Measure for Frequency Coded Data." publised in Kybernetika, 34, 4, 363-368, 1997.
- [6]Jharna Majumdar,Kiran S,"Human Tracking using Particle Filter", International Journal of Computer Applications, Volume 76– No.6, August 2013 .
- [7] Omar Javed and Mubarak Shah," Tracking And Object Classification For Automated Surveillance" Proceeding ECCV '02 Proceedings of the 7th European Conference on Computer Vision-Part IV, 3-540-43748-7, pp. 343-357,2002

BIOGRAPHY



Kritika Gupta Completed B.E in Computer Science from Chhattisgarh Institute of Management Technology Bhilai (C.G.) in 2012 and currently pursuing M.Tech in Nitte Meenakshi Institute of Technology, Bangalore.



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 year