

PERFORMANCE ANALYSIS OF VIDEO SHOT SEGMENTATION ALGORITHM

M. Raja Suguna¹, A. Kalaivani²

¹ Assistant Professor, Department of Computer Science and Engineering, Saveetha School of Engineering, Saveetha Institute of Medical and Technical Sciences, suguna.raj89@gmail.com

² Associate Professor, Department of Computer Science and Engineering, Saveetha School of Engineering, Saveetha Institute of Medical and Technical Sciences, kalaivania@saveetha.com

Abstract:

Video Shot Segmentation is the basic technique in Post Video Processing Application Software. Video Shot Segmentation divides a video into meaningful temporal segments called shots. A shot is composed of series of interrelated images continuously recorded without any interrupt by a single recording device. A video frame is the unique image in the shot. There are two shot boundary change Abrupt and Gradual. The abrupt shot transition is called Cut and gradual shot transition includes dissolve, Fade, Wipe. Identifying the video frame at which shot change occurs is called Shot Boundary Identification. In this paper, a detailed study on Shot Boundary Segmentation methods using conventional Machine learning Techniques and deep learning Techniques are studied.

Key Words

Shot Transition, Cut, Gradual Shot Change, Deep Learning Techniques, Machine Learning Techniques.

1. INTRODUCTION:

A Video has a four-layer Hierarchical Structure. Video is a sequence of non-intersecting units called Scenes. A Scene in a video is a sequence of one or more meaningful segments called Shots. A shot in turn is a sequence of interrelated Images taken continuously by a camera. The single image of a Shot is referred as a Frame.

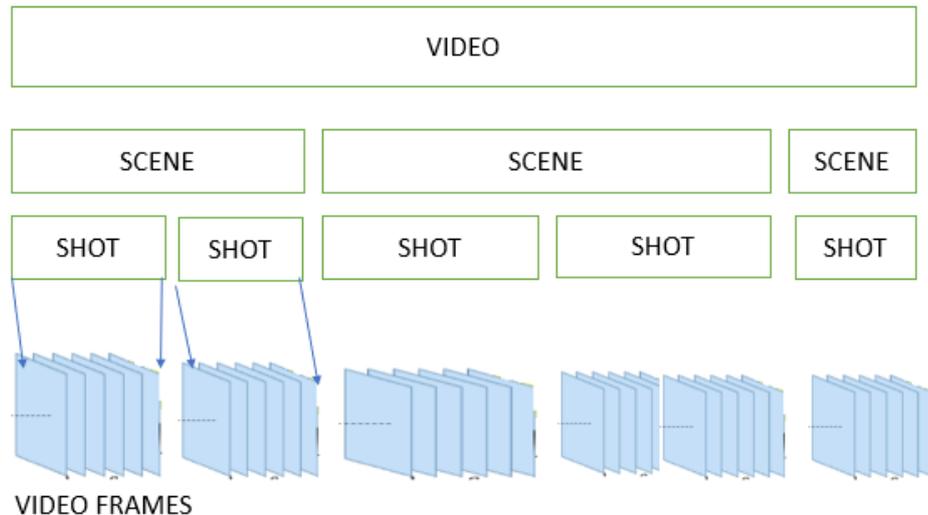


Fig 1. A Video Hierrachy

A Shot is a meaningful temporal segment of a video. Video Shot Boundary Detection Algorithm aims to find out the video frame at which the Shot Change Occurs. A cut is the basic shot transition, a video frame followed by the different or next shot of camera. The second type of Shot transition is Gradual Shot Change, which includes Fade In/Out, Dissolve, Wipe. In a Fade Shot change, the image turns to a one colour generally black or a new image gradually appears on the screen. Fade in shot change is usually appear at the beginning of the shot and fade out appears at the ending of the shot. In a Dissolve the image dissolves to another image, rather than changing into a single colour in Fade, one shot gradually changes to next shot. In Wipe, the pixels of second image get replaced by the pixels of the first image gradually in any of the regular patterns such as iris slow, barn door, star, heart, matrix or in inline from the left edge of the frame.

Most of the works in literature accurately detects the cut shot change than gradual shot Transitions. Many Conventional Machine Learning algorithms and the deep learning techniques have been implemented to detect cut and gradual shot change transitions.

2. RELATED WORKS

The Machine Learning Techniques used for shot boundary identification are pixel difference method, Statistical methods, Histogram Based Methods, Motion Vector Descriptors, Edge Difference, SURF features, SIFT features and deep learning-based algorithms. These methods extract features from each video frame and compared with the threshold value. The adjacent frames showing large differences in feature values and greater than threshold value is identified as the Shot Transition frame. Theoretically all the frames with in a shot will have similar visual content. Thus, when a Cut Shot transition occurs the dissimilarity features between the frames will be large. Hence when gradual Shot transition occurs, these dissimilarity features values will be very small. Other factors that affects the accuracy of shot Change detections are light variation, camera or object motion, similar background, camera option changes such as zooming.

3. SHOT BOUNDARY DETECTION TECHNIQUES

3.1 Pixel Differences:

In this method, the differences between the adjacent frame is calculated by counting the differences in pixel intensity values between two frames when it exceeds the threshold value defined. A shot Change is detected, if the sum of differences in pixel values greater than the second threshold [1]. Zang et al [4] pre-processed the image by using averaging filtering. The averaging filter on the images replaces the pixel value by its average neighbour pixel values. The filter is applied all over the frame. This method is implemented to reduce the noise and effects created by camera motion. They found out the threshold values manually by applying the mentioned pre-processing steps.

Kikukawa [2] used absolute differences of pixel values and compared with the first threshold value and their cumulative pixel difference values compared with second threshold value. Nagasaka [3] used the partial differences of pixel values between the adjacent frames to compare with the first threshold and identifies the pixel change. A Cut is identified when the sum of pixel differences is larger than the second threshold .The pixel differences method is very sensitive to camera motions.

3.2 Statistical Differences:

This method is extension of pixel differences method. Instead of counting the change in pixel intensity values between the frames, statistical measures of grey pixel of the image are considered for comparison. An image is divided into number of regions .The difference in statistical measures of pixels of these regions between the two successive images are calculated for shot boundary detection. The Statistical measures of an image are mean, standard deviation, variance and Median. This method is not very sensitive to noise as in pixel differences method. But the method slows down due to the complexity of coputation Statistical measure's formula [5]. Alattaret al. [6] used statistical method variance for detecting the gradual transition dissolve. Truong et al. [7] detects the gradual Transitions Dissolve, Fade in /out by calculating the statistical measures between the frames. The statistical measure differences of the frames in the start of the Fade out transition is range from very small to zero and vice versa in the case of fade in transition.

3.3 Histogram Based Methods:

A Histogram representation of a Image shows the frequency of pixel intensities in the graphical format. The X-axis consists of pixel intensity value range and the y-axis is consists of frequency of pixel intensity values. Bins in histogram is the frequency of the pixel intensity value. A Colour Histogram shows the distribution of the colour pixel intensity values in an image. This is the most common method for detecting the Shot Boundary detection. The first step is to compute the grey or colour histogram for the two consecutive video frames. If the bit wise difference between the video frames is beyond the threshold defined ,then usually a Cut transition is detected. This method used by researchers in [3] with grey histogram for shot transition detection but it is found to be sensitive to noise, camera and object motion and the flash light variations.

Zang et al [4] compared three techniques such as Pixel variations, Statistical Measure Variation and the Histogram comparison method shot boundary detection. In their Histogram based method, they used two thresholds for the detection of gradual transitions. when the bit wise difference between the two consecutive video frames lies between these two thresholds marks the presence of gradual transition.

First, the grey histogram of all the successive video frames is calculated. Histogram Difference Metric is the sum of the histogram's differences of the successive video frames at given grey level in the given video. The two thresholds namely high threshold and low thresholds which are computed by calculating mean and the standard deviation of the Histogram Difference Metric of all the video frames in a Video. When the histogram difference metric between the two successive video frame is above the high threshold, a Cut is identified. Gradual transitions are detected when histogram difference metric exceeds the low threshold. This method is also called running or adaptive histogram-based approach.

In [9], a region histogram is proposed. A video frame is subdivided into 16 blocks and 64-bin grey histogram is calculated for each region. Histogram Difference metric is calculated for the corresponding regions between two successive video frames. A shot transition is detected if the number of region difference exceeds the count threshold. Lienhart et al. [10] used the twin threshold calculated for RGB Histogram.

Gargi et al. [11] calculated the histogram for different colour spaces and the distances measures. Bin- bin Difference, chi Square test and histogram intersection is used for shot boundary detection. Thounaojam et al. [12] used Histogram difference method using genetic and the fuzzy logic concepts. Their method evaluated by the videos from TRECVID 2001 Video database, which includes gradual shot transition such as dissolve transition and fade transition. Histogram Based Shot Boundary detection is computationally faster than the statistical methods and shows the good trade off between the accuracy and complexity of computation. But it is not more robust to external factors such as camera motion, illumination changes and noise in the videos.

3.4 Edge Based Tracking:

In this method, to determine the shot boundary detection, an edge change calculation is assisted by Edge change ratio and motion vector compensation. It is computationally expensive than the simple histogram-based method and not outperform it. It can be used to remove false positives in both cut and gradual shot change transition in presence flash light occurrences [13]. A shot transition is detected if the position of the edge in the current frame is larger in difference from the positions in the previous frame.

3.5 Motion Based Method:

In this method Motion Vector need to be computed from the compressed domain of video. Block Matching Algorithm used to extract the motion vectors. A video frame divided into number of blocks. The Block Matching Algorithm compares the block in current video frame to each and every block in the successive video frames. A shot transition or a any camera movement will be detected, if the block differences exceeds the defined threshold value. A camera movement within a shot can be considered as gradual transition. The accuracy of shot detection will be decreased with inappropriate selection of motion vectors by block matching algorithm.

Priya et al. [14] extracted motion vector using block matching algorithm and then fused it to other features of the video such as edge tracking, colour spaces and image texture using feature Weighting concept. It detects both abrupt(cut) and the gradual shot transition using Procedural and statistical machine learning methods.

Implementation of shot Transition detection using Motion vector in uncompressed domain of video is computationally more expensive

3.6 Deep Learning Methods:

Implementation of deep learning techniques in computer vision received more attention because of their automatic feature extraction from the images and the videos. Convolutional Neural Network extract the features from the images more rapidly and able to process large number of Images more quickly.

Tonget al. [15] used Convolutional neural Network for shot transition detection. The CNN consists of seven layers and trained with ImageNet database with 1000 classes(tags). The video frames are pre-processed by adaptive method and a group of five video frames is given as input to the CNN. The output is the highest probability tag is given to each video frame. A shot transition is detected when the tags with in the group of five video frames are not similar.

Xu et al. [16] proposed the standard Convolutional Neural Network. The pre-processing is same as that of in [15]. The proposed Convolutional Neural Network is trained with ImageNet dataset of 7000 iterations. After training, the features are extracted from the fully connected FC6 layer .The Cosine Similarity is calculated for the features extracted for each video frame .A Abrupt or cut Transition and Gradual Transitions are measured from these cosine similarity of features extracted from the video frames. Dhinakaran K et. al. discussed the Video Surveillance Wildfire Detection using Dark Convolutional Neural Network [17].

4. Conclusion and Feature Work:

In the digital era enormous quantity of video are generated and processed in the internet every day. An efficient Content based Indexing and retrieval of digital video, Video Techniques for monitoring public, CCTV Footages, Cyber monitoring are needed. A efficient shot boundary detection is basics for the mentioned applications. In this paper, lot of proposed shot boundary detection algorithm using different approaches such as pixel differences,histogram,statistical,motion vector,edges and deep learning methods are reviewed. Each method has its own advantages and disadvantages. The open challenges affecting the accuracy of Shot Boundary detection algorithm are camera movements,change in camera options such as zooming, flash light variations and the same background without variation. Deep Learning Algorithm are able to handle large amount of images can be studied for addressing these challenges.

REFERENCES:

- [1] Lee,M.S,Yang,Automatic Video Parsing using shot boundary detection and camera operation analysis.Pattern Recognition,34(3), 711-719

- [2] Takeshi .k, Kawabuchi,"Development of an automatic Summary Editing System for audio visual resources",Transactions of Instof Electronics,information and communications.1992 ,204-212
- [3] Nagasaka, Akio and Y. Tanaka. "Automatic Video Indexing and Full-Video Search for Object Appearances" VDB ,1991.
- [4] Zhang,Kankanhalli,Smoliar, "Automatic Partitioning of Full-Motion Video Multimedia Systems 1, 10-28,1993,<https://doi.org/10.1007/BF01210504>.
- [5] R. Kasturi and R. Jain, 'Dynamic vision,' in Computer Vision: Principles,IEEE Computer Society Press,Washington ~1991
- [6] Alattar," Detecting and Compressing dissolve regions in Video Sequences with a DVI multimedia Image Compression Algorithm.,1993 IEEE Inter Symp on Circuits and Systems,vol 1,13-16
- [7] Truong,B.T,et al,"New Enhancements to cut ,fade and dissolve detection process in video segmentation", Multimedia ,2000.
- [8] Hanjalic,A."Shot -Boundary Detection:Unraveled and Resolved" IEEE Trans Circuits Video Technology,12,2002,90-105
- [9] John S. Boreczky, Lawrence A. Rowe,"Comparison of VideoShot Boundary Detection Techniques",Proc.SPIE 2670,Storage and Retrieval for Still Image and Video Dbs IV, 13 March,1996;<https://doi.org/10.1117/12.234794>
- [10] Rainer W. Lienhart," Comparison of automatic shot boundary detection algorithms",Proc SPIE 3656, Storage and Retrieval for Image and Video Databases VII,1998
- [11] Gargi, U., R. Kasturi and S. Strayer. ". Performance Characterization of Video-Shot-Change Detection Methods." (2000). IEEE Trans. Circ. 2000, 8, 4761–4766
- [12] Dalton Meitei Thounaojam,T. Khelchandra, M.Singh,S. Roy ,"A Genetic Algorithm and Fuzzy Logic Approach for Video Shot Boundary Detection",ComputationalIntelli and Neuroscience,vol 2016 ,<https://doi.org/10.1155/2016/8469428>
- [13] Wei Heng,King N,"High Accuracy Flashlight Scene Determination for shot boundary Detection, Signal Processing :Image Comuni ,Vol 18,Issue 3, 2003,P 203-219
- [14] Priya, G. and S. Domnic. "Walsh–Hadamard Transform Kernel-Based Feature Vector for Shot Boundary Detection." IEEE Trans on Image Processing 23 (2014): 5187-5197.
- [15] Tong, Wenjing et al. "CNN-based shot boundary detection and video annotation." 2015 IEEE InternSym on Broadband Multi Systems and Broadcasting ,2015:P: 1-5.
- [16] Xu, Jingwei et al. "Shot Boundary Detection Using Convolutional Neural Networks." 2016 Visual Comm and Image Processing ,2016, P: 1-4.

- [17] Dhinakaran. K, R. Anand, et.al (2020)," Video Surveillance Wildfire Detection using Dark Convolutional Neural Network", Test Engineering and Management, Vol.83, pp No. 11601 - 11604.