January-February 2020
ISSN: 0193-4120 Page No. 16321 - 16327

Scene Change Detection in a Television Video

Sumanth S
Assistant Professor, Department of ECE, PESCE
sumanthsmandya@gmail.com
Dr. K. A. Radhakrishna Rao
Professor, Department of ECE, PESCE
karkrao@gmail.com

Article Info
Volume 82
Page Number: 16321 - 16327
Publication Issue:
January-February 2020

Article History
Article Received*: 18 May 2019*
Revised*: 14 July 2019*
Accepted*: 22 December 2019*
Publication*: 28 February 2020*

Abstract:
Scene change detection is very important in video application, including video indexing, semantic features extraction, multimedia information systems, video on demand (VOD) and digital TV. In this work we propose automatic scene change detection algorithm in broadcasted video. The proposed technique uses audio feature - signal energy, energy difference and visual features- frame difference, histogram difference, edge change ratio, face similarity. Experimental results on videos demonstrate better results for scene change detection.

Keywords:*Video on demand, scene change detection (SCD).*

I. INTRODUCTION

TV commercial has become essential business tool of digital media components. Peoples have different Opinion on TV commercial, from viewer point, he may be not interested to watch commercial or while recording he may want to skip the commercial during recording; from businessmen /advertisement owners, they want to verify the broadcasted commercial is as per agreement, it includes the on air time and length of the commercial; from the public institutions, they can monitor whether the broadcaster is following the rules and regulation for the broadcasting the commercial and general program and also they can verify that, during which time the commercial are broadcasted more. However, the variety of TV commercials makes the detection a rather challenging problem.Generally during video browsing, it allows to go for next frame, fast forward can be observed. Usually it is in terms of linear i.e fast forward by 2s, 4s, 16s, so on. But with scene change detection inbuilt it can replace fast forward button with next shot, next scene, etc.General video sequence is shown in figure1. Videos consist of different programs in between programs commercial will be added. Programs consist of scenes; it can be further classified into shots. Shot is a continuous frames acquired by camera without a pause. Shot consist of series of frame.

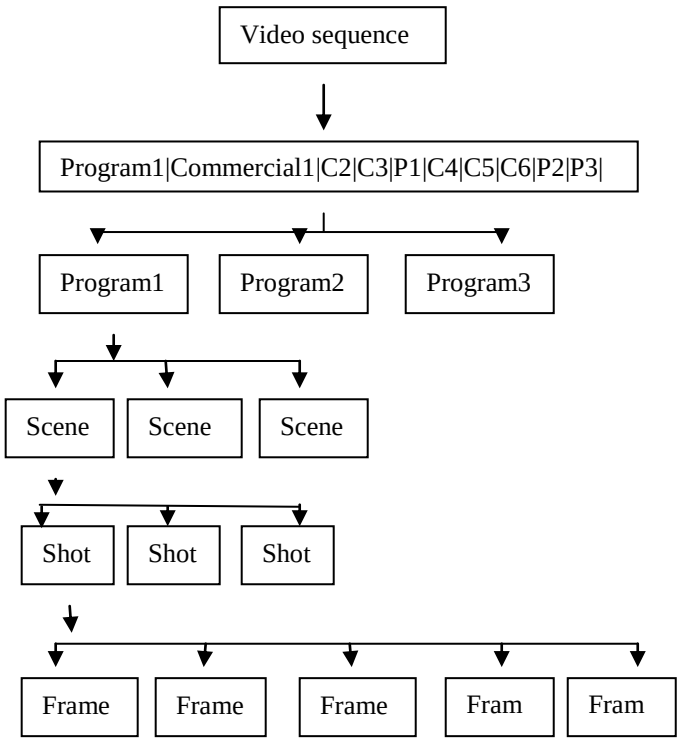


Figure1: General Video Structure

Shots are combined to form scene, in between shots there will be editing like cut, fade, dissolve, wipe, etc. this editing are done to make viewer feel less strain. If the contrast changes quickly between the

January-February 2020
ISSN: 0193-4120 Page No. 16321 - 16327

shots, the viewer eye gets strained quickly. Advertisers to catch audience, they had more number of shots in a scene and makes the editing with special effects. A cut is defined as concatenation of two shot without any film editing. During concatenation of any two shots, there will be film editing like fade, dissolve, wrap over etc.. During a fade, ongoing shot frame intensities gradually decreases this is called as fade out. During a fade_in incoming shot frame intensities increases gradually. If fade_in and fade_out start together then it is called as dissolve [12]. Ongoing shot replaced with another shot from one side to another like a line moves across the screen it is called as wipe.Over a period of time different approaches have been proposed for TV commercial detection problem. In [1] they have used black frames detection and reduction in audio volume but now a day none of broadcaster is adding black frames between the scene transitions. In [2] the proposed algorithm is based TV Logo absence during commercial broadcasting. In [3] they have focused on local multimodal characteristics and global temporal behavior i.e audio class histogram, commercial pallet histogram, text location indicator, scene change rate, blank frame rate.In [4] they have considered successive video shots dependently and finding the temporal coherence between the shots for merging and classification of scene. For classification they have used Hidden Markov Model and Dynamic Bayesian Network. In [5] they have used repetitive use of commercials over time, color difference and audio features. For identifying repeating key frames, similarity metric is used, which is based on the average and standard deviation of HSV values obtained from a 5x5 grid. Generally commercials contain background music while news contains mostly speech. Hence, audio feature as discriminative feature for commercials and News. In [6] to catch audiences during the break, commercial producers polish up their videos as interesting as possible. Videos full of motion and many shot changes. Here they have used two-level commercial detection scheme based on cuts and strong cuts. In [7] they have used to repeated sequence, temporal and chromatic variations with in the clip. In [8] they have used six visual feature- average ECR, Variance ECR, Average Frame Difference, Variance FD, Shot frequency, Black frame Ratio and five audio features-audio break detection (MFCC, Short term energy), speech, music, silence, background. In [9] the first stage exploits six subtitle constraints and an adaptive neurofuzzy interface system model to determine whether a frame contains a subtitle or not. The second uses genetic algorithm. First stage: It comprises a string of letters that is aligned horizontally, it has uniform font and size, all letters are interspaced, it has unique texture and a strong contrast to the background, it is monocolored, it has a fixed style but the literal content is changing regularly. In [10] they have used shot density, Product information Frame, silent sequence and commercial segment length.The rest of the paper is organized as follows. In section II, we introduce a set of features that are used in our algorithm, while detection scheme in section III. In section IV, Experimental results are discussed, section V briefly presents the conclusion.

II. Feature Analysis

The transition of shot can be generally classified as two types. Abrupt shot change (CUT) and gradual transition(GT). The gradual transition can be further classified as fade in/out, dissolve, wipe, etc., here we are using Histogram difference, edge change ratio, AECR, VECR, frame difference, AFD, VFD face detector and signal energy.

Usually Shot boundary detection is based on pixel-based color histogram, edges, motion estimation techniques, etc

A. Histogram Difference

Shot change is usually detected by examining the visual dissimilarity between two consecutive frames [11]. If the dissimilarity is large enough, an abrupt shot change is assumed. Gradual shot change can be detected by accumulating the dissimilarities over successive frame and applying a more sophisticated

January-February 2020
ISSN: 0193-4120 Page No. 16321 - 16327

threshold scheme.The histogram difference between the adjacent frames i and i+1 is defined as

$$E(i) = \sum_{j=0}^{255}\left(H_{i+1}(j) - H_i(j)\right) \quad (1)$$

$H_i(j)$ represents the number of pixels with the gray level j at frame i.

B. Edge Change Ratio

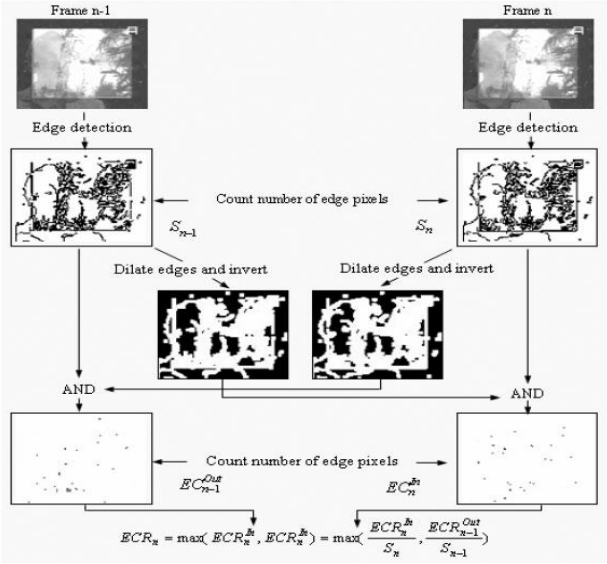


Figure 2:Main steps to compute the edge change fraction

Temporal visual discontinuity usually comes along with structural discontinuity. To measure the dissimilarity between different frames to detect video shot boundary. The edge change ratio (ECR) between $(i-1)^{th}$ frame and i^{th} frame is defined as follows [12],[8]

$$ECR_i = max\left(\frac{X_i^{in}}{S_i}, \frac{X_{i-1}^{out}}{S_{i-1}}\right) \quad (2)$$

The detailed step for Edge change ratio calculation is shown in figure 2.Where ECR_i is the number of edge pixels in frame i, X_i^{in} and X_i^{out} are the number of entering and exiting edge pixels in frame i and i-1, respectively. Average-ECR and Variance-ECR of second 't' are then defined by

$$AECR(t) = \frac{1}{F-1}\sum_{i=1}^{F-1} ECR_i \quad (3)$$

$$VECR(t) = \frac{1}{F-1}\sum_{i=0}^{F-1}(ECR_i - AECR(t))^2 \quad (4)$$

Where, F is the number of frames in one second.

A. Frame Difference

Frame difference(FD) is defined by

$$FD_i = \frac{1}{P}\sum_{k=0}^{P-1}\left|F_k^i - F_k^{i-1}\right| \quad (5)$$

Where, P is the pixel number in one video frame, F_i^m is the intensity value of pixel i in the frame m.

$$\mathrm{AFD(t)} = \frac{1}{\mathrm{F}-1}\sum_{\mathrm{i}=1}^{\mathrm{F}-1}\mathrm{FD_i} \quad (6)$$

$$\mathrm{VFD(t)} = \frac{1}{\mathrm{F}-1}\sum_{\mathrm{i}=0}^{\mathrm{F}-1}(\mathrm{FD_i} - \mathrm{AFD(t)})^2 \quad (7)$$

Where, F is the number of frames in one second.

B. Face Detector

For face detection Viola Jones algorithm have been used [13]. It uses 4 steps, in the first step it extracts haar feature, then it creates an integral image, learning algorithm Adaboost and cascading classifiers. Haar feature uses 4 rectangular window (figure 3). A few properties which are similar to human faces, the eye region is darker than the upper cheeks, the nose bridge region is brighter than the eyes. The matchable facial properties can be in seen in figure 4.

Rectangle features:

Value = Σ (pixels in black area) - Σ (pixels in white area) (8)

It has used two-rectangle features, each feature is related to a special location in the sub-window.Then with the help of learning algorithm Adaboost select the best feature and train the classifier. Cascade architecture identifies the facial or non-facial in all sub windows.

January-February 2020
ISSN: 0193-4120 Page No. 16321 - 16327

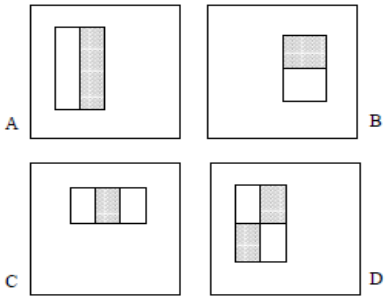
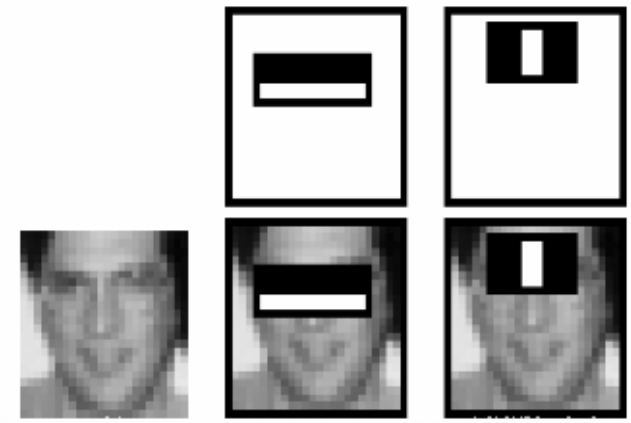


Figure3: Haar features.



Figur 4: Face Detection using Haar cascade.

C. Signal Energy

During the scene change audio volume will reduce and has smooth transition. Generally average short term energy of TV commercial is higher than other type of video.The short term energy is given by

$$STE(m) = \frac{1}{N}\sum_{n=0}^{N-1}[x(n)h(w-n)]^2 \quad (10)$$

N is the number of sample in each audio frame. Here each audio frame is divided into 20 ms. m is audio frame number.

$$h(n) = \begin{cases} 1 & 0 \le n \le N-1 \\ 0 & otherwise \end{cases}$$

Average short-term energy

$$avgSTE = \frac{1}{SampleNum}\sum_{m=1}^{SampleNum} STE(m) \quad (11)$$

Where, SampleNum is number of frame in one second

III. SCENE CHANGE DETECTION

Based on the features discussed in section 2, A multi feature scene change detection system is designed, which is shown in figure 5.

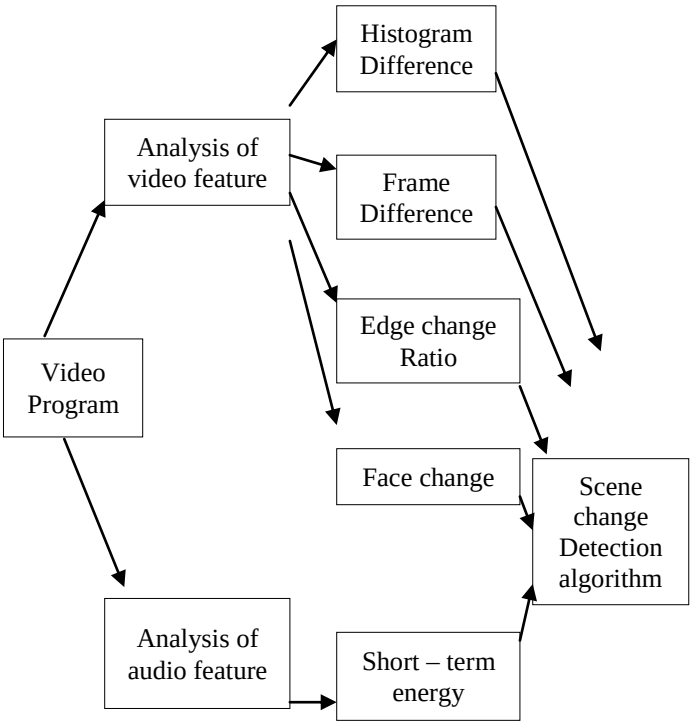


Figure 5: Scene Change Detection

Here the recorded TV program is taken as input. The video program as video along with audio embedded to it. Firstly the audio and video information are separated and analysis of video and audio are done separately. Video features are extracted from the frames. Frame classification is based on the compression standard, but for audio we have made 25ms each frame. The features are extracted from both audio and video frame as discussed in section II. Face change as extra feature in the scene detection system improves the performance. We have extracted the face in each video frame. The mean and standard deviation of the detected face have been saved. The similarities between the faces are analyzed over the length of window. If the similarity is more it belongs to same scene else it is treated as scene change. It is going help in avoiding the false detection which may occur due to other features.

The face detector algorithm works as follows

1. Faces are extracted in each frame using viola jones algorithm. i.e$\boldsymbol{f_i^t}$, where, i is the frame number. l is the number of faces in frame i.
2. For each face mean and standard deviation is calculated.

January-February 2020
ISSN: 0193-4120 Page No. 16321 - 16327

$$f\_m_i^l = \frac{1}{N}\sum_{j=1}^{N} f_i^l(j) \quad (12)$$

$$f\_s_i^l = \sqrt{\frac{1}{N}\sum_{j=1}^{N}\left[f_i^l(j) - f\_m_i^l\right]^2} \quad (13)$$

3. Find the difference between the faces in consecutive frame.

$$f\_m\_d_i^l = \sum_{i=1}^{N-1} f\_m_i^l - f\_m_{i+1}^l \quad (14)$$

$$f\_s\_d_i^l = \sum_{i=1}^{N-1} f\_s_i^l - f\_s_{i+1}^l \quad (15)$$

4. See the similarity in the face-mean-difference (f_m_d) and face-std-difference (f_s_d) over length of frames.
5. Set the threshold to detect the face change over the frame.
6. If the difference is less than the threshold the frame belongs to the previous scene else it is a scene change.

The histogram difference eliminates the object motion in a video. Frame difference gives the motion in a video. Edge change ratio helps in the detection of fade in, fade out, dissolve, cut. Here adaptive threshold is used for scene change detection. Audio feature short term energy helps in identification of silence sequence detection. Usually during the scene change 2 to 3 frame of audio signal have very low signal energy.

IV. EXPERIMENTATION AND RESULT

In order test the algorithm, recorded Television program from SaralJeevan channel (kannada) is used. Video programs were split into smaller video segment for experimentation. Each video segment include program as well as commercials. For testing 10 different video clips were used. The features discussed in section II are extracted individually from each clip. The algorithm is developed and simulated on matlab software2018a. The table 1 shown the scene point detected from individual feature. The individual feature as given more number of data points from the threshold taken, but in the table it has shown that the nearest value to the scene changes. Unwanted points from individual feature are removed with the collaboration of other feature.Actual the face doesn't change at the scene point only, it may require some more frame to change. No faces are also treated as feature for scene change detection. Here by setting the threshold for individual feature and finding the similarity with other features, correct scene change points are detected.

Table 1 : Scene Points Detected from video clip1

Scene change frame number	Histo-gram Diffe-rence	Frame Diffe-rence	Edge change Ratio	Face Change	Short Term Energy
1110	1106	1106	1112	1102	1115
1680	1688	1688	1684	1681	1684
2640	2646	2647	2644	2645	2645
3690	3691	3691	3691	3696	3692
4650	4654	4654	4652	4657	4647

Table 2 : Scene Points Detected from video clip2

Scene change frame number	Histo-gram Diffe-rence	Frame Diffe-rence	Edge change Ratio	Face Change	Short Term Energy
1260	1266	1267	1261	1245	1262

January-February 2020
ISSN: 0193-4120 Page No. 16321 - 16327

2130	2131	2131	2133	2087	2135
3180	3186	3185	3186	3369	3184
4140	4148	4150	4144	4142	4146
5460	5468	5478	5463	5466	5622

Table 3 : Scene Points Detected from video clip3

Scene change frame number	Histo-gram Diffe-rence	Frame Diffe-rence	Edge change Ratio	Face Change	Short Term Energy
1650	1659	1659	1653	1652	1727
3360	3361	3362	3362	3366	3362
4290	4294	4294	4295	4294	4495
5640	5645	5645	5646	5645	5645
7140	7142	7143	7143	7148	7142

We have simulated the algorithm for all dataset available with us. The simulation result shows that all the scene change points are detected in addition that false positive also detected. The features have given change point according to the shot change in a scene. There may be more number of persons in a scene, we look for at least one match in the specified window. The scene change detection point precision rate.Overall average frame difference for accurate detection on different feature is as follows: for all 10 clips available

Table 4 : Difference in scene change Frame Number identification for features

Histogram Difference	Frame Difference	Edge change Ratio	Face Change	Short Term Energy
5	7	6	4	5

For error calculation it is taken interms of frame number.In the dataset, we have collected there are 47 scene points. Through simulation it was able to accurately detect 42 points. 5 were unable to detect. In future paper, we will modify the threshold for individual features in algorithm and additional features will be used.

V. CONCLUSION

This paper presents scene change detection algorithm with the fusion of audio and visual features. The highlight of this paper is on face change detection algorithm, which other paper have not concentrated. It has helped for better classification of scene change points. The results were given good result. The main future work will focus on the audio features to detect the temporal change of music and classify the program and commercial segment separately.

REFERENCES

[1] David A sadlier, Sean Marlow, Noel O'Connor, Noel Murphy "Automatic TV advertisement detection from MPEG Bitstream", The journal of the pattern recognition society,2002

January-February 2020
ISSN: 0193-4120 Page No. 16321 - 16327

[2] Alberto Albiol, Maria Jose Ch. Fulla, Antonio Albiol "Detection of TV Commercials" , IEEE, ICASSP 2004.

[3] Masami Mizutani, ShahramEdadollahi, Shih-Fu Chang " Commercial Detection in HetrogenousVideo Streams Using Fused Mutli-Modal and Temporal Features" IEEE, ICASSP 2005.

[4] Tie-Yan Liu, Tao Qin, Hong-Jiang Zhang "Time Constriant Boost For TV Commercials Detection" IEEE, International Conference on Image Processing(ICIP) 2004.

[5] Pinar Duygulu, Ming-yu Chen, Alexander Hauptmann "Comparison and Combination of Two Novel Commercial Detection Methods" IEEE International Conference on Multimedia and Expo(ICME) 2004.

[6] Jen –HaoYeh, Jun-Cheng Chen, Jin-HauKuo, Ja-Ling Wu "TV Commercial Detection in News Program Videos" IEEE 2005.

[7] John M. Gauch, AbhishekShivadas "Identification of New Commercials using Repeated Video Sequence Detection" IEEE 2005.

[8] Xian-Sheng Hua, Lie Lu, Hong-Jiang Zhang "Robust Learning –Based TV Commercial Detection" IEEE 2005.

[9] Yo-Ping Huang, Liang-Wei Hsu, Frode-EikaSandens "An Intelligent Subtitle Detection Model for Locating Television Commercials" IEEE Transaction on System, Man and Cybernetics –Part B:Cybernetics, Vol.37 No.2, April 2007.

[10] Xiang Wang, ZongmingGuo "A Novel Real-time Commercial Detection Scheme" IEEE The 3rd international conference on Innovative Computing Information and Control 2008.

[11] Li Meng, Yong Cai, Min Wang, Yuanxing Li "TV Commercial Detection Based on Shot Change and Text Extraction" IEEE 2009.

[12] R. Zabih, J. Miller, K. Mai, "A Feature-Based Algorithm for Detecting and Classifying Scene Breaks," Proc of ACM Multimedia 95, San Francisco, CA, pp. 189-200, Nov. 1995.

[13] Paul Viola, Michael Jones "Robust Real-time Object Detection" Second International Workshop on Statistical and Computational Theories of Vision– Modeling, Learning, Computing, and Sampling Vancouver, Canada, July 13, 2001.

[14] Lie Lu, Hao Jiang, HongJiang Zhang, "A Robust Audio Classification and Segmentation Method" 9thACM Multimedia, pp. 203-211, 2001.