ISSN (Print) : 0974-6846
Indian Journal of Science and Technology, *Vol 9(7), DOI: 10.17485/ijst/2016/v9i7/84338, February 2016* ISSN (Online) : 0974-5645

Video Shot Boundary Detection using Gray Level Cooccurrence Matrix

Dalton Meitei Thounaojam[1,2*], Sudipta Roy[2] and Kh. Manglem Singh[3]

[1]Department of Computer Science & Engineering, National Institute of Technology, Silchar - 788010, Assam, India; dalton.meitei@gmail.com

[2]Department of Computer Science & Engineering, Assam University, Silchar - 788011, Assam, India; sudipta.it@gmail.com

[3]Department of Computer Science & Engineering, National Institute of Technology, Imphal – 795001, Manipur, India; manglem@gmail.com

Abstract

Objectives: The objective of this paper is to find out the abrupt transitions between consecutive shots in a video with less false detection and high F1 score. **Method/Analysis:** This paper presents a video shot boundary detection approach using Gray Level Cooccurrence Matrix (GLCM). The proposed system can roughly be divided into feature extraction using GLCM and the application of the abrupt shot boundary detection. In the first step, the frames are converted into gray level and GLCM is calculated from each frame in the video. Secondly, correlation coefficient is calculated from the GLCM of two consecutive frames of the video. A threshold is set to identify the shot boundaries of the video. The proposed system can detect abrupt transitions effectively with less false detection in the uncompressed domain. **Findings:** The proposed system can able to achieve an average F1 score of 93.51%, which is achieve due to the reduced false detection. **Novelty/ Improvement:** The proposed system uses the GLCM matrix directly instead of calculating the contrast, entropy,etc, i.e., the proposed system is purely based on the correlation of the pixel's co-occurrence. The proposed system also reduces the false detection thereby increasing the precision and F1 score.

Keywords: Abrupt, Gradual, Gray Level Cooccurrence Matrix, Shot Boundary Detection,Video Segmentation

1. Introduction

With the advances in the Internet system and the growth of social media, online tutorial and lectures, online shopping and online business system the generation of multimedia content on the web increases tremendously which leads to a problem of proper indexing and retrieval of video. The multimedia contents mainly include videos, images, sounds, text and other interactive information. For proper and effective retrieval of the multimedia contents, an effective indexing and retrieval tools are necessary. Large multimedia data like video need to be taken care because a single video contains much information unlike image since video is a collection of images (frames) ordered in a sequential and meaningful manner. In a content based video indexing and retrieval system, temporal video segmentation technique forms the first stage, where the video is divided into meaningful segments called shots. After finding out the shots of a video, key frames or features are extracted for proper indexing and then used for retrieval process. A detailed explanation including advantages and disadvantages of the current techniques of image and video retrieval is provided in Suguna et al.[1]

In temporal video segmentation the main task is to find out the shot boundaries – abrupt and gradual transitions[2]. Abrupt transitions are caused by camera off and on. Gradual transitions are caused by editing effects. The types of gradual transitions are fade-in, fade-out, dissolve and wipe.

Researchers have tried to detect the shot boundaries using the video content features such as sound, text, frames and motions3. The most commonly used feature for temporal segmentation is the histogram4. The histogram is calculated using gray-level and/or color

* *Author for correspondence*

space like RGB[5–7], HSV[8,9] or Lab10 of a frame. The fast and simplest methods for shot boundary detection is the gray or color histogram comparison[8,11] between two frames and the use of SVD[5] in the frames histogram matrix. In Jadon et al.[7], a fuzzy classification is applied to the color histogram difference. In Küçüktunç et al[10], a Fuzzy color histogram using the L*a*b color feature is proposed to find out the shot boundary and later for content-based copy detection system. In Uma and Ramakrishnan[12], edge and edge histogram are calculated from frames using Non Subsampled Shearlet transform for sports video classification.

In Lu and Shi [8], a frame skipping technique[13] and an inverted triangular pattern is proposed for shot boundary detection. In Tong et al.[13], Convolutional Neural Network model is used to extract TAG's from the image for shot boundary detection.

In Ralph et al.[14], many parameters like histogram difference, edge, MPEG and pixel difference are considered for shot boundary detection. Image features like pixel distance[15], discrete wavelet coefficient [16,17] and edge counts are also used for video segmentation. Vila et al.[18] proposed Tsallis mutual information and Jensen–Tsallis divergence for shot boundary detection where in both cases use Tsallis entropy.

In our method, we propose to calculate the Gray Level Co-occurrence Matrix of each frame and find out the correlation coefficient between two frames of the video. A local thresholding technique is used to find out the transitions.

Section 2 and 3, we discussed about Gray Level Cooccurrence Matrix and Correlation Coefficient. Section IV shows the system design to find abrupt transition. Sections V give the experimental results followed by conclusion in Section VI.

2. Gray Level Cooccurrence Matrix

Grey Level Cooccurrence Matrix is also called as **Grey Tone Spatial Dependency Matrix**[19–22]. It is used to find out the texture feature of an image. GLCM is a two dimensional matrix which is computed using a displacement vector *d*, and orientation θ. *d* values ranging from 1 to 10 and every pixel has eight neighboring pixels with θ value 0°, 45°, 90°, 135°, 180°, 225°, or 315° respectively.

GLCM calculates how often a pixel with gray-level value *i*occurs horizontally adjacent to a pixel with the value *j in an image I*Given an $M \times N$ neighbourhood of an input image containing *G* gray levels from 0 to *G* - 1, the elements of the GLCM are given by Equation 1 [21].

$$P(i,j \mid d,\theta) = WQ(i,j \mid d,\theta) \tag{1}$$

where, $W = \dfrac{1}{(M-d)(N-\theta)}$

$$Q(i,j \mid d,\theta) = \sum_{n=1}^{N-\theta}\sum_{m=1}^{M-d} A$$

$$A = \begin{cases} 1 & \text{if } f(m,n)=i \quad \& \quad f(m+d,n+\theta)=j \\ 0 & \text{otherwise} \end{cases}$$

and *f(m,n)*is the intensity at sample *m* and line *n* of the neighbourhood.

3. Correlation Coefficient

Correlation coefficient (ρ_{XY}) is used to measures the strength and the direction of a linear relationship between two values. It is computed using Equation 2 [7].

$$\rho_{XY} = \frac{n\sum_{p=1}^{n}\sum_{q=1}^{n} X_p Y_q - \sum_{p=1}^{n} X_p \sum_{q=1}^{n} Y_q}{\sigma_X \sigma_Y} \tag{2}$$

where, n is the number of elements in the frame.σ_X and σ_Y are the means of frames X_p and Y_q respectively and are given by

$$\sigma_X = \sqrt{n\sum_{p=1}^{n} X_p^2 - \left(\sum_{p=1}^{n} X_p\right)^2} \tag{3}$$

$$\sigma_Y = \sqrt{n\sum_{q=1}^{n} Y_q^2 - \left(\sum_{q=1}^{n} Y_q\right)^2} \tag{4}$$

4. Proposed System

The proposed system includes a preprocessing system, where the each frame is converted from JPEG image to gray image. Gray Level Cooccurrence Matrix is computed for each matrix and similarity between consecutive frames is found out using Pearson's correlation coefficient. The proposed system is given in Algorithm 1. An experimental threshold is taken to classify the possible abrupt transition as given in Equation 5.

After classifying the possible abrupt transition frames

from the frames without transitions, Algorithm 2 is used to find the final abrupt transition.

Algorithm 1: Shot Boundary Detection algorithm using GLCM

Input : Video

Output : Shot boundary

V ← Read the video into frames;

for k := length of the video – 1

G1 ← calculate GLCM of V_k th frame;

G2 ← calculate GLCM of V_{k+1} th frame;

C← correlation(G1,G2) ;

if $C \leq T$

then $t_k = 1$;

Else

$t_k = 0$;

End

End

Abrupt ← apply Algorithm 2

To detect the transitions, a threshold is selected and a transition detector, $TD = \{ t_1, t_2, t_3,...,t_k\}$, , where, k is the length of ρ_{XY}. The value of t_k is determined by Equation 5.

$$t_k = \begin{cases} 0 & \text{if } \rho_{XY} > T \\ 1 & \text{otherwise} \end{cases} \quad (5)$$

where, T is a predefined threshold and it is determined through experiment.

The transition detector is not sufficient to determine the type of transition (in this case abrupt transition only). So, a simple algorithm for abrupt transition detection is applied as shown below:

Algorithm 2: Abrupt transition detection algorithm

Input : Transition detector, $TD = \{ t_1, t_2, t_3,...,t_k\}$

Output : Abrupt transition

for k = length of TD

if $t_k = 1$ & t_k-$1 = 0$ & $t_{k+1} = 0$ & $t_{k-2} = 0$ & $t_{k+2} = 0$

then declare an abrupt transition between frame k and $k+1$;

End

End

5. Experimental Results

TRECVid 2007 video test dataset are used for our experimentation. TRECVid 2007 data consist of documentary videos of various lengths. All the video data provided are in MPEG compressed and can be downloaded from NIST after request of data. Also one video from OPEN VIDEO PROJECT and a documentary video on "**A Wild Dog's Tale**" are also used for system evaluation. The description of the video used in our experimentation is given in Table 1.

Table 1. Description of the video and Shot Boundary detected using GLCM

Video	Frame size	Frame no.	Abrupt
A Wild Dog's Tale.mp4	720 × 1280	1-4000	94
hcil2001_01.mpg	240 × 352	670-4066	22
BG_3097.mpg	288 × 352	1-10000	41
BG_11369.mpg	288 × 352	1-3700	21
BG_35111.mpg	288 × 352	1-10000	28

In the proposed system, the GLCM is calculated for each of the frames in a video and the correlations between the consecutive frames are measured using Pearson's correlations. Figure 1 shows an example of the correlation coefficient and the sequence number of the first frames where the correlation is calculated.

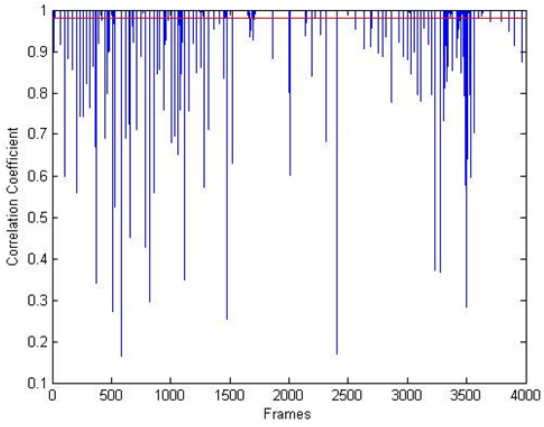

Figure 1. Shows the correlation coefficient between two frames of the GLCM matrix for the video A Wild Dog's Tale.

On experimentation, it is observed that the threshold lies in the ranges [0.9, 0.98] which give an overall performance, F1 score, of 80% to 93% by using the video shown in Table 1. Throughout our experiment, 0.98 is taken as the threshold. It is also to be noted that the whole process is perform using GLCM only without finding out certain properties like the contrast, entropy, correlation and homogeneity of a frame.

In Table 2, the comparison is done between the SVD[5] and our proposed system and our proposed system shows better

Table 2. Proposed system performance

Video	Proposed system			SVD [5]		
	Recall %	Precision %	F1 score	Recall %	Precision %	F1 score
A Wild Dog's Tale.mp4	94.68	93.68	94.17	96.80	85.71	90.16
hcil2001_01.mpg	81.81	94.73	87.79	100	61.11	75.86
BG_3097.mpg	92.68	95	93.82	100	77.35	87.22
BG_11369.mpg	95.23	95.23	95.23	95.23	68.96	79.99
BG_35111.mpg	100	93.33	96.54	100	73.68	84.84

performance. The main reason behind the comparison with [5] is that histogram and SVD are used by many researchers for shot boundary detection [4].

Three parameters *recall, precision* and *F1 score* are chosen to evaluate the performance of the detection.

The equations for recall, precision and F1 score are as follows:

$$\text{Recall} = \frac{N_C}{N_C + N_M} \times 100 \quad (5)$$

$$\text{Precision} = \frac{N_C}{N_C + N_F} \times 100 \quad (6)$$

$$\text{F1} = 2 \times \frac{\text{Recall} \times \text{Precision}}{\text{Recall} + \text{Precision}} \quad (7)$$

where, N_C is the number of correct detections, N_M is the number of missed detections, and N_F is the number of false detections.

Figure 2 and 3 shows some sample of the falsely detected and correctly abrupt transitions by our proposed system from the sample video.

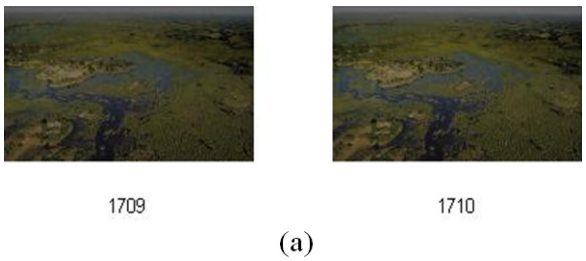

Figure 2. **(a)** Shows the false detection for the video "A Wild Dog's Tale".

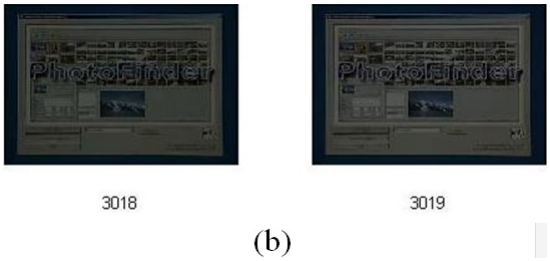

Figure 2. **(b)** for the "hcil2001_01" respectively.

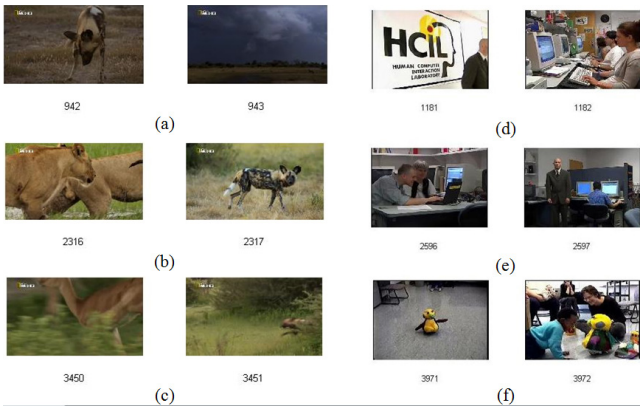

Figure 3. **(a)**, **(b)** & **(c)** the abrupt transitions for the video "A Wild Dog's Tale" and **(d)**, **(e)** & **(f)** for the "hcil2001_01" respectively.

5. Conclusion and Future Works

This paper proposes a video shot boundary detection approach using Gray Level Co-occurrence Matrix. In the proposed system, the frames are converted into gray level image as the preprocessing step and GLCM is calculated from each frame in the video. Similarity between consecutive frames in a video is calculated using Pearson's correlation coefficient from the GLCM of the two frames. A threshold is set to identify the possible shot boundaries of the video. The proposed system detects abrupt transitions effectively, in terms of F1 score, as compared to an existing system.

The system can able to detect gradual transitions but it is very sensitive to intensity of the frame which results in too much false detection of gradual transition. So, overcoming this problem is the future work. The future work also includes the key frame or key feature detection and scene classification using GLCM. GLCM can be used to find contrast, entropy, correlation and homogeneity of a frame. This can be used to find out the key frames or key features of shots/scenes.

6. Acknowledgement

Sound and Vision video is copyrighted. The Sound and Vision video used in this work is provided solely for research purposes through the TREC Video Information Retrieval Evaluation Project Collection.

7. References

1. Suguna S, Kumar CR, Jeyarani DS. State of the art: a summary of semantic image and video retrieval techniques. Indian Journal of Science and Technology. 2015 Dec; 8(35):1–12. doi: 10.17485/ijst/2015/v8i35/77061.
2. Koprinska I, Carrato S. Temporal video segmentation: a survey. Signal Processing: Image Communication. 2001; 16(5):477–500.
3. Thounaojam, DM, Amit T, Singh KM, Roy S. A survey on video segmentation. Intelligent Computing, Networking, and Informatics, AISC. 2014; 243:903–912.
4. Smeaton AF, Over P, Doherty AR. Video shot boundary detection: seven years of TRECVid activity. Computer Vision and Image Understanding. 2010; 114(4):411–18.
5. Yihong G, Liu X. Video shot segmentation and classification. International Conference on Pattern Recognition; 2000. p. 860–3.
6. Zuzana C, Kotropoulos C, Pitas I. Video shot segmentation using singular value decomposition. International Conference on Acoustics, Speech, and Signal Processing; 2003. p. 181–4.
7. Jadon RS, Chaudhury S, Biswas KK. A fuzzy theoretic approach for video segmentation using syntactic features. Pattern Recognition Letters. 2001; 22(13):1359–69.
8. Hameed A. A novel framework of shot boundary detection for uncompressed videos. International Conference on Emerging Technologies. 2009; 274–9.
9. Lu ZM, Shi Y. Fast video shot boundary detection based on SVD and pattern matching., IEEE Transactions on Image Processing. 2013; 22(12):5136–45.
10. Küçüktunç O, Güdükbay U, Ulusoy O. Fuzzy color histogram-based video segmentation. Computer Vision and Image Understanding. 2010; 114(1):125–35.
11. Zhang HJ, Kankanhalli A, Smoliar SW. Automatic partitioning of full-motion video. Multimedia Systems. 1993; 1(1):10–28.
12. Maheswari SU, Ramakrishnan R. Sports video classification using multi scale framework and nearest neighbor classifier. Indian Journal of Science and Technology. 2015 Mar; 8(6):529–35. doi:10.17485/ijst/2015/v8i6/61067.
13. Tong W, Song L, Yang X, Qu H, Xie R. CNN-based shot boundary detection and video annotation. IEEE International Symposium on Broadband Multimedia Systems and Broadcasting. 2015: 1–5. doi: 10.1109/BMSB.2015.7177222.
14. Ralph FM, Craig R, Daniel T, Michael G. Metrics for shot boundary detection in digital video sequences. Journal Multimedia Systems. 2000; 8(1):37–46.
15. Sun J, Wan Y. A novel metric for efficient video shot boundary detection. Visual Communications and Image Processing Conference. 2014: 45–8. doi: 10.1109/VCIP.2014.7051500.
16. Parthasarathy MB, Srinivasan B. Increased security in image cryptography using wavelet transforms. Indian Journal of Science and Technology. 2015 Jun; 8(12):2–8. doi:10.17485/ijst/2015/v8i12/62433.
17. Venkateswaran S, Desai UB. DWT based hierarchical video segmentation. IEEE International Symposium on Circuits and Systems. ISCAS 2002; 2002. p. 815–8. doi: 10.1109/ISCAS.2002.1010349.
18. Vila M, Bardera A, Xu Q, Feixas M, Sbert M. Tsallis entropy-based information measures for shot boundary detection and keyframe selection. Signal, Image and Video Processing. 2013; 7(3): 507–20.
19. Haralick RM, Shanmugam K, Dinstein Its'Hak. Textural features for image classification. IEEE Transactions on Systems, Man and Cybernetics. 1973; 3(6):610–21.
20. Mokji MM, Bakar ASAR. Gray level co-occurrence matrix computation based on haar wavelet. Computer Graphics, Imaging and Visualisation. 2007: 273–9.
21. Albregsten F. Statistical texture measures computed from gray level cooccurrence matrices. Technical Note, Department of Informatics; University of Oslo: Norway; 1995
22. Baraldi A, Parmiggiani F. An investigation of the textural characteristics associated with gray level cooccurrence matrix statistical parameters. IEEE Transactions on Geoscience and Remote Sensing. 1995; 33(2):293–304.