A Shot Boundary Detection Technique Based On Local Color Moments In YC_BC_R Color Space

S.A.Angadi[1] and Vilas Naik[2]

[1]Department of Computer Science Engineering,
Basaveshwar Engineering College,Bagalkot
vinay_angadi@yahoo.com
[2] Department of Computer Science Engineering,
Basaveshwar Engineering College,Bagalkot
vilasnaik_h@rediffmail.com

ABSTRACT

Color moments are measures that can be used to differentiate images based on their color. Once computed, these moments provide a measure for color similarity between images. This color similarity can become a measure of finding difference between consecutive frames in a video. Finding the difference between temporally arranged frames in video is one of the issues in shot boundary detection which will segment the video in to meaningful units called shots. So a Shot boundary detection method based on primitives of color moments is proposed in this paper. The YCbCr color model is used for the experimentation. Each frame of the video is partitioned in to 8 X 8 smaller sub blocks. The first three color moments of every sub blocks of each frame in the video are computed. The feature vector is formed from the moments and compared between successive frames. The difference value higher than a predefined threshold at any pair of consecutive frames indicates the presence of a shot boundary between frames. By using YCbCr color space, the influence of illumination change and shadows is reduced. Experiments have been conducted on the standard soccer video database to evaluate the performance of the proposed method.

KEYWORDS

Shot boundary, Color moments, Adaptive thresholding.

1. INTRODUCTION

Shot detection is used to segment a video sequence into basic temporal units called shots which is a series of interrelated consecutive pictures taken contiguously by a single camera and representing a continuous action in time and space. It is commonly the first step in the process of indexing, characterizing and retrieving of video. The principle methodology of shot boundary detection is to extract one or more features from every frame of a video sequence, compute the difference of features for consecutive frames and comparison of these differences to a given threshold. Each time the threshold is exceeded, a shot boundary is detected. The various approaches differ in the type of features used. The detection of boundaries between video shots provides a basis for almost all of the existing video segmentation and abstraction methods [1]. However, it is quite difficult to give a precise definition of a video shot transition since many factors such as camera motions may change the video content significantly. Usually, a shot is

defined to be a sequence of frames that appears to be continuously captured by the same camera [2].

It is observed that a shot can generally encompass camera motions such as pans, tilts, or zooms, and video editing effects such as fades, dissolves, wipes [3]. In particular, video shot transitions are usually grouped into two types: abrupt/sharp shot transitions, which is also called cuts, where a frame from one shot is followed by a frame from a different shot, and gradual shot transitions, such as cross dissolves, fade-ins and fade-outs, and various other editing effects respectively.

Several Models have been proposed for Shot boundary detection modeled on statistical properties of image pixels, compression algorithms, edge differences or Histogram differences. The Histogram differences method with two thresholds is used in order to detect gradual transitions[4]. The Audio and motion features have been used to improve shot boundary detection. Classification of audio according to silence, speech, music, or noise and use this information to verify shot boundaries hypothesized by image-based features[5]. An algorithm for shot clustering using speaker identification is presented in[6]. The use of motion features alone or with histogram differences to improve the shot boundary detection is discussed in [7]. The method in [8] combines motion features with pixel differences. Although there exist many algorithms for shot detection, this work is an attempt to devise computationally efficient shot detection model that is capable of detecting boundaries of shots for segmenting the video into collection of meaningful units called *shots*. The proposed model is based upon color moments in YCbCr color space. It is observed in [9] that the quantization effects of the color histogram is eliminated and the color distribution is characterized by its moments. This motivated to investigate the application of color moments for shot boundary detection(SBD).

The proposed technique uses first two moments the mean and variance at individual layers of YCbCr color model in block of NXN size for frames in video. The moments are employed in constituting feature vector further used for finding similarity between consecutive frames to construct continuity signal. The temporal position in frame index where continuity measure cross beyond predefined threshold indicates shot boundary.

The rest of the paper is organized into 4 sections, Section 2 elaborates about color moments and their use in SBD. The proposed model is presented in section 3. The experimental results and analysis is presented in section 4. The section 5 brings out the conclusion of the work.

2. COLOR MOMENTS FOR SHOT BOUNDARY DETECTION

Color moments are measures that can be used to differentiate images based on their features of color. Once calculated, these moments provide a measurement for color similarity between images. These values of similarity can then be compared to the values of images indexed in a database for tasks like image retrieval. The basis of color moments lays in the assumption that the distribution of color in an image can be interpreted as a probability distribution. Probability distributions are characterized by a number of unique moments (e.g. Normal distributions are differentiated by their mean and variance). It therefore follows that if the color in an image follows a certain probability distribution, the moments of that distribution can then be used as features to identify that image based on color.

The basis of color moments lays in the assumption that the distribution of color in an image can be interpreted as a probability distribution. Probability distributions are characterized by a number of unique moments . It therefore follows that if the color in an image follows a certain probability distribution, the moments of that distribution can then be used as features to identify that image

based on color[9]. The three central moments of an image that represents the color distribution are Mean, Standard deviation and Skewness.

The work is restricted to **YCbCr** color space representing color by three values, Y′ is the Luminance component and C_B and C_R are the blue-difference and red-difference chroma components. Moments are calculated for each of these channels in an image. Therefore A frame of video is characterized by 9 moments 3 moments for each 3 color channels. We will define the ith color channel at the jth image pixel as p_{ij}. The three color moments can then be defined as:

Mean

$$E_i = \sum_{N}^{j=1} \frac{1}{N} p_{ij} \tag{1}$$

Mean can be understood as the average color value in the image.

Standard Deviation

$$\sigma_i = \sqrt{(1/N \sum_{N}^{j=1} (p_{ij-} E_i)^2)} \tag{2}$$

The standard deviation is the square root of the variance of the distribution.

Skewness

$$s_i = \sqrt[3]{(1/N \sum_{N}^{j=1} (p_{ij-} E_i)^3)} \tag{3}$$

Skewness can be understood as a measure of the degree of asymmetry in the distribution

A function of the dissimilarity between color distributions of two consecutive frames from video can be defined as the sum of the weighted differences between the moments of the two distributions. Formally this is:

$$d_{mom}(H, I) = \sum (E^1_i - E^2_i) + (\sigma^1_i - \sigma^2_i) + (s^1_i - s^2_i) \tag{4}$$

Where

(H, I): are the two image distributions being compared

i : is the current channel index (e.g. 1 = Y, 2 = Cr, 3 = Cb)

E^1_i, E^2_i :are the first moments (mean) of the two image distributions

σ^1_i, σ^2_i :are the second moments (std) of the two image distributions

s^1_i, s^2_i :are the third moments (skewness) of the two image distributions

In the proposed work this dissimilarity value is measure of discontinuity against frame index. If the dissimilarity measure is higher than a predefined threshold indicates the presence of a Shot boundary.

3. PROPOSED METHODOLOGY

In this section, proposed model devised using color moments is presented . Given a video containing many frames, compute the color moments for each frame and the Euclidean distance measure is used to measure the similarity between the frames. Based on the preset threshold, a shot is said to be detected if the dissimilarity between the frames is above the threshold. It is a well known fact that the color distributions of a frame is represented by its color moments and probability distribution is uniquely characterized by its moments, we compute the color moments of a frame to capture the characteristics of a frame and construct a feature vector for each frame as discussed in the following.

Feature Extraction

Color moments of an image in the Y(Luminance), Cb (Chrominance of blue), Cr (Chrominance of Red) color space are very simple to calculate. Color moments are very effective for color-based image analysis. They are especially important for classification of images based on color, image similarity for retrieval mechanism. Color moments to be calculated are in fact, statistical moments. Generally the first order (mean of color values) ,the second order moments (standard deviation of color values) and third moment skewness have been used in this work.

Algorithm

Step1: *Every frame extracted from video sequence to be converted first into Y,Cb, and Cr format from the raw R,G,B format.*

Step2 : *An image has to be partitioned into sub-blocks. Deciding on optimal number of sub-blocks is a qualitative question and has to be decided as per size of image . Optimal choice used in this work is 8x8 as shown in figure 1*
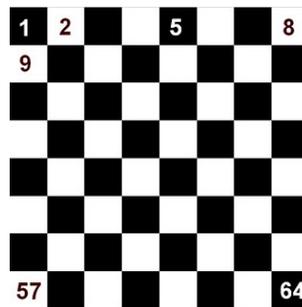

Figure 1.0: 8x8 sub blocks scanning while calculating color moments

Step 3: *By scanning each sub-blocks as indicated in Figure. 1, calculate for each sub-block in Y, Cb, and Cr planes three measures: statistical mean and Standard deviation and skewness.*

In the proposed work   320*320 frame has been partitioned into 8*8 sub-blocks. Each sub-block size is then 40*40 pixels. Color moments for this blocks are nothing but the statistical mean and standard deviation and skewness of this 40*40 pixel values. For total of 64 blocks. There are three planes Y, Cb, and Cr. This way total number of subblocks is 64*3 = 192. For each sub block, calculate one mean value and one standard deviation and skewness value, thus total of 192*3 = 576 values; These 576 values formatted in 64x 9 matrix form make the feature vector for each frame. Each row of 9 values represent one sub block. The each row of feature vector is represented as

$$F_n = [\, E_{yn} \;\; \sigma_{yn} \;\; s_{yn} \;\; E_{cbn} \;\; \sigma_{cbn} \;\; s_{yn} \;\; E_{crn} \;\; \sigma_{crn} \;\; s_{crn} \,] \qquad (5)$$

Where

$E_{yn} \;\; \sigma_{yn} \;\; s_{yn}$: Mean, standard deviation and skewness values of Y component

$E_{cbn} \;\; \sigma_{cbn} \;\; s_{yn}$: Mean, standard deviation and skewness values of C_b component

$E_{crn} \;\; \sigma_{crn} \;\; s_{crn}$: Mean, standard deviation and skewness values of C_r component

The function of the similarity between color distributions of two consecutive frames from video can be defined as the sum of the Euclidian differences between the feature vector of moments values of the two frames.

Dissimilarity function

The dissimilarity between frame f_i and f_{i+1} dm is computed as sum of the absolute differences between elements from feature vectors for f_i and f_{i+1} or the Euclidian distance between feature vector of f_i and feature vector of f_{i+1}

$$dm(\, f_i , f_{i+1}) = \sum_{i=0}^{n-1} \parallel F_i - F_{i+1} \parallel \qquad \text{(Absolute difference)} \qquad (6)$$

$$dm(\, f_i , f_{i+1}) = \sqrt{\sum (F(j,k) - F_{i+1}(j,k))} \qquad \text{(Euclidian distance)} \qquad (7)$$

Where

F_i and F_{i+1}: feature vector containing sub block wise moments of Y Cb and Cr channels of frames.

Detection of Shot boundary

The proposed model is based on color moments. Given a video which contains many frames, the color moments for each frame are computed and the Euclidean distance measure is used to measure the similarity between the frames. Based on the set threshold, a shot is said to be detected if the dissimilarity between the frames is higher than the threshold value.

Temporal statistics using color moments of frames are computed. Once the color moments discussed in Section II are extracted as per the steps discussed in section III from the current frame, they will be compared with the color moment values of next frame. A shot boundary is detected when the feature difference dm calculated as per equation (7) shows sharp change greater than predefined threshold determined by an adaptive scheme. The continuity signal which is graph of frame difference against frame index on time axis is shown in the Figure 2.

dm
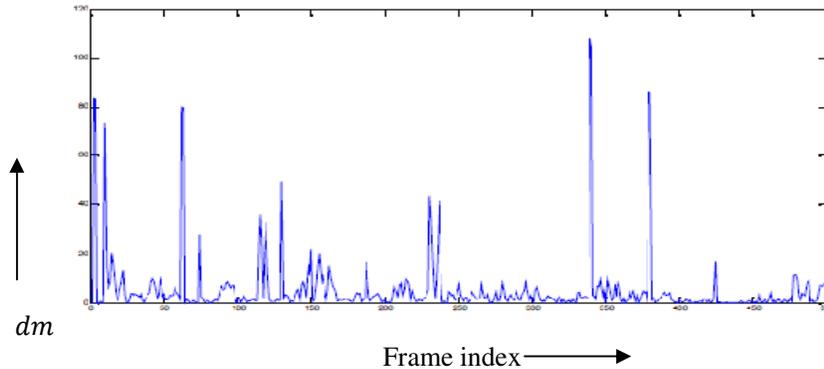
Frame index⟶

Figure 2.0: plot of dissimilarity value plotted against frame index
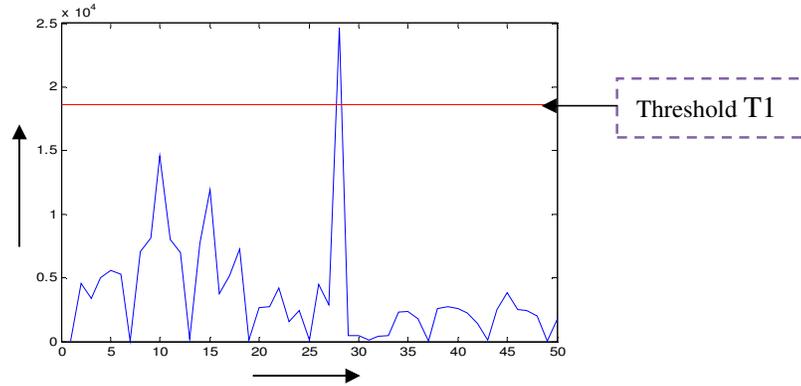
Figure 3.0: Threshold value on frame difference graph.

Threshold Selection

The problem of choosing the appropriate threshold is a key issue in   the shot detection algorithms. Heuristically chosen global thresholds are inappropriate because experiments have shown that the threshold for determining a segment boundary varies from one shot to another which must be based on the distribution of the frame-to-frame differences of shots. So adaptive threshold is more reasonable than global threshold. In proposed algorithm,   local window based threshold calculation method is verified . A window of m frames is built and then the mean value μ is calculated and the threshold T is defined as

$$T= 5\times (\text{Mean } \mu \text{ of the widow size m}) \qquad (8)$$

Where value of m is decided as 50 empirically. The dissimilarity value greater than threshold T1 indicates shot boundary. The figure 3 shows shot boundary being detected between frame 28 and frame 29 in the experimental data set.

4. EXPERIMENTAL RESULTS

This section presents the results of the experiments conducted to corroborate the success of the proposed model. The experimentation is conducted on set of YOU TUBE soccer and cricket videos and UCIF sports data set. All experiments are performed on a Core2 Duo 2.40 GHz Windows machine with 2GB of RAM. The performance of the proposed model is evaluated using precision and recall as evaluation metrics. The precision measure is defined as the ratio of number of correctly detected cuts to the sum of correctly detected and falsely detected cuts of a video data and recall is defined as the ratio of number of detected cuts to the sum of detected and undetected cuts. These parameters were obtained for the proposed model on eight different video samples.

$$Recall = \frac{number\ of\ correctly\ detected\ boundaries}{number\ of\ true\ boundaries}$$

$$Precision = \frac{number\ of\ correctly\ detected\ boundaries}{number\ of\ totally\ detected\ boundaries},$$

$$(9)$$

The results for 03 test videos randomly selected from UCF and You tube action data is presented in the following table. The Figure 4.0 is the plot of Plot of frame dissimilarity measure for the cricket video. The Figure 5.0 is the plot of threshold level on Plot of frame dissimilarity measure and indication of the shot cut/s in the video.

Table 1.0: Precision and Recall metrics for the proposed algorithm

	Size	No. of frames tested	Shot Cuts detection performance of Proposed Algorithm	
			Recall	Precision
Video 1	2.0 MB	800	1.00	0.900
Video 2	1.5 MB	500	1.00	0.870
Video 3	2.5 MB	1000	0.857	0.950
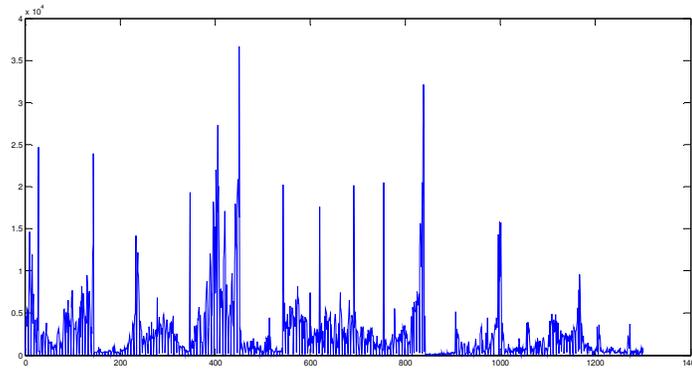
Figure 4.0 Plot of frame dissimilarity measure for the cricket video
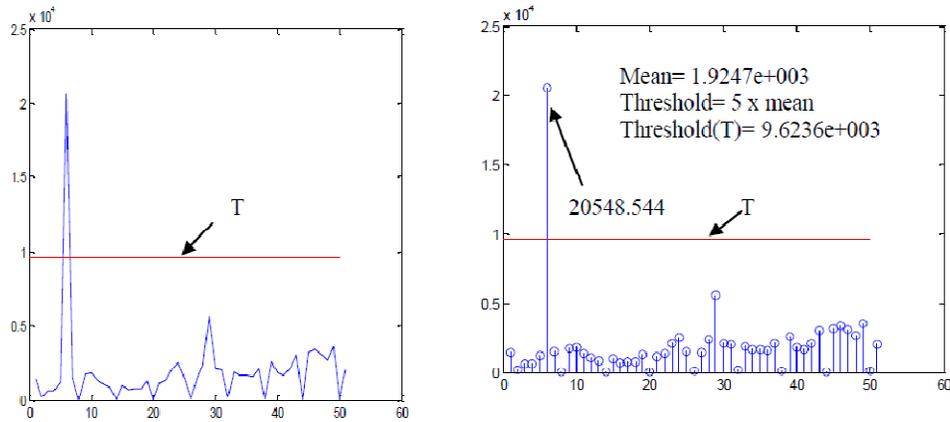
Figure 5.0 plot of threshold level on Plot of frame dissimilarity measure

Frame 753   Frame 754   Frame 755   Frame 756   Frame 757   Frame 758

Figure 6.0 Detection of shot boundary between frame 755 and frame 756

5. CONCLUSION

An efficient texture features based model for identifying shot boundaries of a video segment is proposed by considering the significant merits of color moments.The information conveyed by every color moments computed for an image in YCbCr Color Space is used to find difference between two frames in a video. The difference between consecutive frames to detect shot transitions is computed as Euclidian distance between feature vector containing color moments mean, standard deviation and skewness of Y(Luminance), Cb (Chrominance of blue), Cr(Chrominance of red) values of frame. The shot boundary is detected wherever difference value is more than predefined threshold formulated using adaptive scheme. Experimental results on standard UCIF and YOU TUBE video data reveal that the proposed model is robust and detects shot boundaries efficiently.

REFERENCES

[1]   Hanjalic, A. Shot-Boundary Detection: Unraveled and Resolved? IEEE Transactions on Circuits and Systems for Video Technology, Vol. 12(2), pp. 90-105 (2002).

[2]   Gargi, U., R. Kasturi, and S. H. Strayer, Performance Characterization of Video-Shot-Change Detection Methods, IEEE Transactions on Circuits and Systems for Video Technology, Vol. 10(1), pp. 1-13 (2000).

[3]   Ziyou Xiong,Regunathan Radhakrishnan,Ajay Divakaran,Yong Rui, and Thomas S. Huang "Unified Frame work for video summarization", Elsevier Academic Press.

[4]   Zhang, H., A. Kankanhalli, and S. Smoliar, Automatic Partitioning of Video, Multimedia Systems, Vol. 1(1), pp. 10-28.

[5]   Saraceno, C. and Leonardi, R., "Audio as a Support to Scene Change Detection and Characterization of Video Sequences", Proc. Int. Conf. Acoustics, Speech, and Signal Processing, Munich, Germany, pp. 2597-2600.

[6]   Nam, J., Cetin, E., and Tewfik, A. "Speaker Identification and Video Analysis for Hierarchical Video Shot Classification", Proc. Int. Conf. Image Processing, Santa Barbara,CA, October, (1997).

[7]   Phillips, M. and Wolf, W., "Video Segmentation Techniques for News," in Multimedia Storage and Archiving Systems, C.-C. Jay Kuo, Editor, Proc. SPIE 2916, 243-251.

[8]   Shahraray, B., "Scene Change Detection and Content-Based Sampling of Video Sequences", in Digital Video Compression: Algorithms and Technologies, Rodriguez, Safranek, Delp, Eds., Proc. SPIE 2419, pp. 2-13.

[9]   Shih, J.-L.; Chen, L.-H.. "Colour image retrieval based on primitives of color moments," Vision, Image and Signal Processing, IEEE Proceedings, vol.149, no.6, pp. 370- 376, Dec (2002).

[10] M. Stricker and M. Orengo. Similarity of color images. In In SPIE Conference on Storage and Retrieval for Image and Video Databases III, volume 2420, pages 381392, Feb. 1995.

Authors

S. A. Angadi is a Professor in the Department of Computer Science and Engineering, Basaveshwar Engineering College, Bagalkot. He earned a Bachelor Degree in Electronics and Communication Engineering from Karnataka University and a Master Degree in Computer Engineering from the Sri Jayachamarajendra College of Engineering, Mysore and a PhD in Computer science from the Department of Studies in Computer science, University of Mysore, India. He has completed many research and consulting projects of AICTE, India, under RPS,MODROBS ,TAPTEC schemes. His research areas are Image Processing and Pattern Recognition, Character Recognition, Fuzzy Systems, Neural Networks and Genetic Algorithms, Optimization and Graph Theoretic techniques, Embedded Systems, Intelligent Systems and Web Technology. He is a co-auther of a book on C-programming language. He is life member of professional bodies like IEEE,ISTE and IETE.

Vilas Naik received BE(Electronics and Communication) from Karnataka   University Dharwad and Master of Engineering in Computer Technology from Shri Guru Govind Singh College of Engineering Nanded under Sri Ramanand Teerth Marathawada University Nanded . He is currently a research scholar registered to Visvesvaraya Technological University, Belgaum in the area of Image and Video processing working on the issues of Multimodal Video summarization and selection. Currently he is working as Assistant Professor in the Department of Computer science and Engineering, Basaveshwar Engineering College Bagalkot. His subjects of interest are Image and Video processing, Data Communications and Computer Networks, Co mputer Architectures and Multimedia computation and Communication.