Genre based Video Indexing and Retrieval using Visual Features

Tushar Ratanpara[1] Dr. K. H. Wandra[2] Dr. N. M. Patel[3]
[1]Student [2]Dean [3]Associate Professor
[1,2,3]Department of Computer Engineering
[1,2]C. U. Shah University, Wadhwan City, Gujarat, India [3]BVM Engineering College, Vallabh Vidhyanagar, Gujarat India

Abstract— With recent advances in digital video coding and transmission, there is a strong demand for a procedure that index, retrieve and classify multimedia data according to users need. In this work, an approach is implemented for genre based video song indexing & retrieval using visual clues. Song sequences are classified into different genre like dance, romantic and sad. Visual features such as shot duration and actor movement is computed for genre based video song sequence classification. Approach is tested on movies of various genres. Accuracy obtain for song genre classification is 75.45 %.

Key words: Genre based Video Retrieval, Genre based Video Indexing

I. INTRODUCTION

Large amount of multimedia data is increasing every day. Lots of time is required to find desired data from large amount of multimedia data. Therefor there is a requirement of procedure that organize, index, retrieve and classify multimedia data according to users need. Traditionally, textual based retrieval method is used to find desired data. In text based retrieval users have to manually assign text or metadata to images and videos [1] [2] [3] [9] [11] [12]. When large amount of images and videos are available in database, huge amount of human work is required in manual annotation of image/video. This manual annotation of image/video is time consuming process. There is a another limitation in text based method like annotation of image and video depends on user interpretation means same image may perceived differently by different people. Content based video indexing and retrieval is used to overcome above limitations.

Bollywood movie contains action scenes, comedy scenes and video songs sequences. Song sequences will be classified into different genre like dance, romantic and sad based on visual features like shot duration and actor movement. Dance an classic songs will be separated based on shot duration because the shot duration in dance songs mostly between 2 to 5 seconds and in classic songs it is greater than 10 seconds [4] [14]. Classic songs will be further divided into romantic and sad songs based on actor movement. In romantic songs actor perform some light dance steps therefor movement is present in shots as shown in figure 1. However in sad songs actor moves hardly in a shot [4] which is shown in figure 2.

Song Genre	Shot Duration	Actor movement
Dance	Low	High
Romantic	High	Medium
Sad	High	Low

Table 1: Visual features of song genre

As shown in table 1, shot duration in dance song is low as compare to romantic and sad song. Actor movement in dance song, romantic song and sad song is high, medium and low respectively.

Fig. 1: Actor movements in romantic song

Fig. 2: Actor movements in sad song

II. RELATED WORK

Audio features based genre classification includes the classification of audio signals into different genre like speech, song, music, etc. Audio features are classified into timber, rhythm, pitch, tonality and dynamics. Y. Zhu, Z. Ming and Q. Huang [5] have proposed a method on clip-based support vector machine to classify audio signals into six classes like music, pure speech, speech with music, silence, speech with environmental sound. Various features like zero crossing rate (ZCR), harmonic ratio, bandwidth, silence ratio, spectrum flux, sub-band energy, low short time energy ratio and high ZCR ratio are extracted from each frame. The mean and standard deviation is calculated for one audio clip to get clip based feature. SVM classifier is used to recognize audio types and 93.1% accuracy is achieved.

George and Perry [6] have presented work on hierarchical music genre classification. Three sets of features are used to represent rhythmic content, pitch content and timbral texture. Timbral texture features includes spectral centroid, spectral roll off, spectral flux, time domain zero crossings, MFCC, analysis and texture window and low energy feature. Beat histogram is used for visual representation so that anyone can easily differentiate genre of songs. Pitch content feature is based on multiple pitch detection techniques. The numbers of standard statistical pattern recognition (SPR) classifiers are used for classification purpose. Classifiers used in this paper are Simple Gaussian, Gaussian mixture model and K- nearest neighbor. Music is further divided into ten different genres like classical, country, hip-hop, rock, disco, reggae, jazz, pop, blues and metal. The classification accuracy has been achieved is 61%.

Visual features are used to classify video into different genre like movies, news, sports, song,

advertisement, serial, etc. S. Jain and R. Jadon have proposed work on movie genre classifier using neural network [7]. Only audio information may not be sufficient for movie clip genre identification so here audio and visual features are used for movie clip genre classification. Visual features extracted from movie clips are motion computation, shot length, lighting key and color dominance. Shot segmentation is done by pixel difference and histogram intersection. Feed forward neural network with back propagation learning algorithm is used as classifier.

Audio features are not sufficient to identify video song sequence genres from the Indian Hindi music because there is a large difference in composition of western and Hindi music. For example, the drum is used in all genres of an Indian Hindi music whereas in western music, it is mostly found in pop music.

The rest of the paper is organized as follows: Section 3 describes proposed approach. Experimental setup is given in Section 4. Experimental results are explained in Section 5 followed by conclusion in Section 6.

III. PROPOSED APPROACH

Abstract model of our proposed approach is shown in figure 3. Initially, video song sequence indexes are extracted using method given in [8] [13]. It is given as an input to the system.

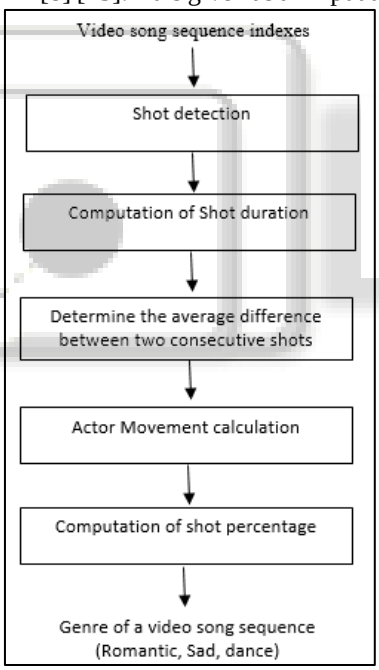

Fig. 3: Abstract model

A. Computation of Shot Duration

For shot duration calculation, shot boundaries in a video should be determined. Histogram based method will be used to detect shot boundaries in a video. Color histogram intersection in HSV color model is used. Each histogram consists 16 bins which are eight component of hue, four component of saturation and four component of value. Intersection of histogram is expressed as equation (1).

$$d(i) = \sum_{n=1}^{16} \min(H_i(n), H_{i+1}(n)) \quad (1)$$

Where H_i and H_{i+1} represent the histogram of i^{th} and $i+1^{th}$ frames. n varies from 1 to 16. 16 is total number of bins in each histogram. d(i) represent the intersection of histogram H_i and H_{i+1}. A frame i will be considered as a shot boundary if its d(i) value is less than threshold value [10]. Threshold value is calculated as below

$$\text{Threshold} = \mu - \alpha\sigma \quad (2)$$

Where α=3, μ is average of intersection of histogram vector (d) and σ is standard deviation of d. n is size of histogram intersection vector (d).

$$\mu = \frac{\sum_{i=1}^{n} d(i)}{n} \quad (3)$$

$$\sigma = \sqrt{\frac{\sum_{i=1}^{n} (d(i)-\mu)^2}{n}} \quad (4)$$

Shot duration will be calculated as the difference between start time and stop time of a shot. For song S, average shot duration $AV_{sd}(S)$ is defined as,

$$AV_{sd}(S) = \frac{\sum_{k=1}^{n} D_k}{n} \quad (5)$$

Where n is number of shots in a song and D_k is duration of k^{th} shot. Dance and classic songs within a movie will be separated based on $AV_{sd}(S)$

$$\text{Genre}(S) = \begin{cases} \text{Dance} & \text{if}(AV_{sd}(S) < t_1) \\ \text{Classic} & \text{if}(AV_{sd}(S) \geq t_1) \end{cases} \quad (6)$$

Where t1 is threshold value computed as,

$$t_1 = \mu_1 - \beta\sigma_1 \quad (7)$$

Where β =1/2, μ_1 is average of all $AV_{sd}(S)$ in a movie and σ_1 is standard deviation of all $AV_{sd}(S)$ in a movie.

$$\mu_1 = \frac{\sum_{S=1}^{T} AV_{sd}(S)}{T}$$

$$\sigma_1 = \sqrt{\frac{\sum_{S=1}^{T} (AV_{sd}(S)-\mu_1)^2}{T}} \quad (8)$$

Where T is number of song in a movie.

B. Computation of Actor Movement

Actor movement shows movement in a video. Classic songs will be classified into two sub-genres like romantic and sad based on actor movement. For actor movement calculation, every image from all shots of a song will be taken for image based comparison. The difference between two consecutive images imagei and imagej is defined as,

$$D = ABS(image_i - image_j) \quad (9)$$

The average difference between all images in a shot is defined as,

$$M = \frac{\sum_{i=1}^{n} D_i}{n} \quad (10)$$

Where M is actor movement in a shot and n is number of images in a shot. Average actor movement in a song S is defined as,

$$AV_{am}(S) = \frac{\sum_{i=1}^{t} M(i)}{t} \quad (11)$$

Where M is actor movement in a shot and t is number of shot in a song. For song sequence index 2 from movie1, actor movement is 391340 with 56 shots.

Classic song S will be categorized into two sub-genre as below,

$$\text{Genre}(S) = \begin{cases} \text{Sad + Romantic} & \text{if}(AV_{am}(S) \geq t_2) \\ \text{Sad} & \text{if}(AV_{am}(S) < t_2) \end{cases} \quad (12)$$

Where $AV_{am}(S)$ is average actor movement in a song and t_2 is threshold value.

$$t_2 = \mu_2 - \gamma \quad (13)$$

Where γ =100000 and μ_2 is average of $AV_{am}(S)$ of classic song.

$$\mu_2 = \frac{\sum_{S=1}^{t} AV_{am}(S)}{t} \quad (14)$$

Where t is number of classic songs in a movie. Sad sequences may not be classified by using actor movement when there is more movement in a song. Sad songs can be

found during break-up of hero-heroine and funeral scenes in a movie. When sad song shows some event which occurs before break up or funeral in a movie then actor movement in a song is high. At that time shot percentage with shot duration greater or equal to 6 seconds will be calculated.

C. Computation of Shot Percentage

Number of shot with shot length greater or equal to 6 seconds will be calculated from remaining sad and romantic sequences. Shot Percentage with shot duration greater or equal to 6 seconds is computed as

$$P = (S * 100) / t\_shot$$

Where P is shot percentage, S is number of shot with length greater or equal to 6 seconds and t_shot is total number of shots in a song. Sad + romantic sequences will be categorized as below,

$$\text{Genre}(S) = \begin{cases} \text{Sad} & \text{if}(P > 22) \\ \text{Romantic} & \text{if}(P \leq 22) \end{cases} \quad (15)$$

IV. EXPERIMENTAL RESULTS

A. Implementation Platform Details

Proposed approach is implemented and tested on platform given below.

1) Hardware Specification

– Processor: Intel Core i3-4005U CPU 1.70 GHz
– RAM: 4 GB

2) Software Specification

– OS : Windows 7 Ultimate
– System type: 64 bit OS
– Front End: MATLAB R2013a

B. Tools and Technology

This section present the tools and technology used for implementing the method that have been mentioned in earlier chapters. Entire work is implemented in MATLAB R2013a.

C. Dataset Design

Video song sequences are extracted using [8]. Video song sequences are having genres like sad, romantic and dance are used in experiment. Proposed approach is tested on 111 video song sequences.

Table 2 shows number of songs, correctly identified song and accuracy for each genre. Accuracy obtain for song genre classification using visual features is 75.45%.

Genre	No. of Songs	Correctly Identified Song	Accuracy (%)
Dance	33	26	78.78
Romantic	62	45	72.58
Sad	16	12	75.00
Total	111	65	75.45

Table 2: Experimental results

Table 3 shows song genre identification results using audio features such as spectral flux, ZCR, MFCC, spectral rolloff and spectral centroid on SVM classifier [4].

Genre	No. of Songs	Correctly Identified Song	Accuracy (%)
Pop	27	22	81.48
Romantic	27	18	66.67
Tragic	27	19	70.37
Total	81	59	72.84

Table 3: Genre identification results using audio features [4]

Figure 4 shows comparison of song genre identification using audio and visual features.

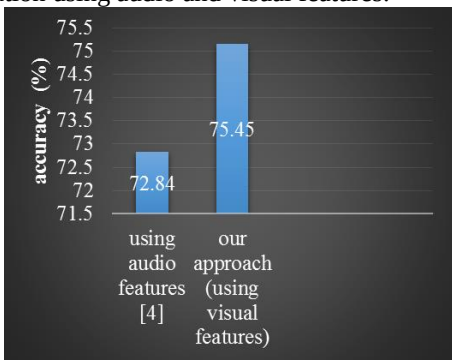

Fig. 4: Comparison of our approach

V. CONCLUSION

Genre based video indexing and retrieval system is proposed. Visual features like shot duration and actor movement are computed. Video song sequences are classified in romantic, sad and dance genre. Experimental results are carried out using various songs of different movies. Visual features give higher accuracy than audio features. Overall, accuracy is achieved is 75.45.

REFERENCES

[1] http://www.ukessays.com/essays/psychology/limitations-of-text-based-image-retrieval-psychology-essay.php
[2] Cees G. and Marcel Worring, "Multimodal Video Indexing: A Review of the State-of-the-art", Springer Multimedia Tools and Application, Vol. 25, pp. 5-35, 2005.
[3] http://musicians.about.com/od/glossary/g/Music-Genre.htm
[4] Doudpota Muhammad, Sumanta Guha and Junaid Baber (2013), "Mining movies for song sequences with video based music genre identification system", Information Processing & Management, pp. 529-544.
[5] Y. Zhu, Z. Ming and Q. Huang, "Automatic audio genre classification based on support vector machine", IEEE Third International Conference on Natural Computation, vol. 1, pp. 517-521, 2007.
[6] George Tzanetakis and Perry Cook, "Musical Genre Classification of Audio Signals", IEEE Transaction on Speech and Audio Processing, Vol.10, pp. 293-302, 2002.
[7] Sanjay Jain and R. Jadon, "Movies Genres Classifier using Neural Network", IEEE Computer and Information Sciences, pp. 575-580, 2009.
[8] T. Ratanpara and N. Patel "Protagonist and deuteragonist based video indexing and retrieval system for movie and video song sequences", International Journal of Advanced Intelligence Paradigms, Inderscience Publisher, 2017.
[9] T. Ratanpara and N. Patel "Singer Identification Using MFCC and LPC Coefficients from Indian Video Songs", Emerging ICT for Bridging the Future - Proceedings of the 49th Annual Convention of the Computer Society of India (CSI) Volume 1, Vol. 337, pp. 275-282, 2015.
[10] Hui-Yu Huang, Weir-Sheng Shih and Wen-Hsing Hsu, "Movie Classification using Visual Effect Features", IEEE Signal Processing Systems, pp. 295-300, 2007.

[11] Chandni Dhamsania, and T. Ratanpara "A survey on Human action recognition from videos." International Conference on Green Engineering and Technologies (IC-GET), pp. 1-5, 2016.
[12] Chandni Dhamsania, and T. Ratanpara. "Human Action Recognition Using Trajectory-Based Spatiotemporal Descriptors." Proceedings of the 5th International Conference on Frontiers in Intelligent Computing: Theory and Applications, Springer, pp. 1-9, 2017.
[13] Darji Mittal, Narendra Patel, and Zankhana Shah. "Extraction of video songs from movies using audio features." International Symposium on Advanced Computing and Communication (ISACC), 2015, pp. 60-64, 2015.
[14] Darji Mittal, Narendra Patel, and Zankhana H. Shah. "A Review on Audio Features based Extraction of Songs from Movies, International Journal of Advance Engineering and Research Development 2015.

