Scientific Journal of Impact Factor (SJIF): 5.71

e-ISSN (O): 2348-4470
p-ISSN (P): 2348-6406

International Journal of Advance Engineering and Research Development

Volume 5, Issue 04, April -2018

CONTENT BASED INDEXING AND VIDEO SONG SEQUENCES RETRIEVAL SYSTEM

Dr. K. H. Wandra

Principal, GMP Polytechnic, Rajula

Abstract — *a movie has colossal amount of data, which includes dialogs, songs and actions. It is very difficult to index each frame according to its content. To mitigate it, there is a need to develop content based indexing and retrieval system. This paper is designed to index video song sequences from the movie. Firstly, segmentation is performed to devise segments of the movie. For each segment, audio channel data is computed. Histogram is plotted. Root mean square is calculated and used to find out range of video song sequences. Simulation results are carried using movies of different genres.*

Keywords-*Content based image retrieval, video indexing, image retrieval, song sequence, amplitude*

I. INTRODUCTION

Video is made from multiple scenes. A scene is made from multiple shots. A shot is made from multiple frames. Video is at first layer while frames are at last layer. Longer video means high number of frames at last level. Video processing is a particular case of signal processing, that often employs video filters and where the input and output signals are video streams or video files [1] [2] [3]. Video indexing, it is the process of providing viewers a way to navigate and access contents easily similar to book indexing. The selection of indexes derived from the content of the video to help organize metadata and video data [4] [5] [6]. Video browsing is the interactive process of skimming through video content in order to interactively check if the video content is relevant or to satisfy needed information. Video browsing also known as exploratory video search [7] [8].

There are significant research work has been done in the area of content based video indexing and retrieval. A video file contains three dimensions of data portions 1) visual 2) audio and 3) text. According to Brezeale and Cook [9] video classification has been classified in three approaches such as visual based approaches, audio based approaches, text based approaches. Generally, combination of these approaches successfully applied over a period of time by many authors; out of it combination of audio-video approaches gives better result. Also classification is applied in different manner. Some authors took video as a whole entity such as action movies, comedy movies, romantic movies, while some authors classify specific feature from video like action scene and comedy scene from Bollywood Hindi movie. There is a so many effort has been done based on cinematic principles or concepts from Bollywood Hindi movie theory [9]. For example, comparison between comedy movies vs. horror movies. Horror movies scenes contains low lighting levels as compare to comedy movies scenes. Therefore, horror movies has higher number of dark scenes rather than comedy scenes. Action movies scene motion more faster rather than romantic movies. Using this kind of cinematic principles yield to very accurate results in classification. However, using only cinematic principles is not sufficient for classification. Cinematic principles can be easily applied to audio as well. Audio is available is in term of channels. Every video contains audio channel. When audio is integrated with scenes it gives more feeling to that scene. Therefore, specific type of audio are integrated in with video give more feeling such as horror scene audio effect is noisy, romantic scene audio effect is soft.

K. Subashini, S. Palanivel, V. Ramaligam [10] have presented work on video genre classification like news, sports, movies, advertisement and serial. MFCC and color histogram are used as audio and visual feature respectively. Support vector machine (SVM) is used for classification and segmentation. 39 features are used for each audio signal, 13 from the length of parameterized static vector, 13 delta coefficients and acceleration coefficients. 64 bin histogram is used for color image. Video frame rate is 25 and audio sample rate is 8 kHz.

The rest of the paper is organized as follows. Proposed embedding and extraction algorithms are explained in section II. Experimental results are presented in section III. Concluding remarks are given in section IV.

II. PROPOSED ALGORITHM

Figure 1 represent abstract model of our proposed system. Initially, Movie is taken as an input, which is partitioned into various segments. Amplitude value is computed from each segment. Threshold is identified using root mean square, which leads to identification of video song sequences.

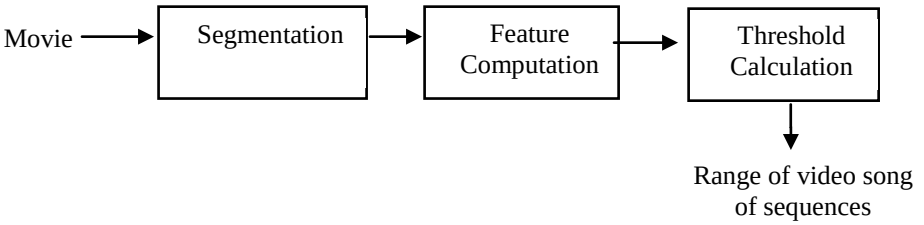

Figure 1. Abstract model of proposed system

A. *Segmentation*

To find out segments from movie, 10 seconds audio channel data is extracted from movie and stored in an array. Similarly, task is performed on whole audio channel of a movie. Array of 1 * X is created, where X is number of segments in a movie. For example, if movie length is 2 hours then size of array is 1 * 720. Now, Feature computation is performed on each segments, which is discussed in next section.

B. *Feature Computation*

Amplitude value is computed for each segment. It is stored in an array. Equation 1 is used to calculate it.

$$s_k = s_k + \begin{cases} 1 & if\ X(i) > 0 \\ 0 & otherwise \end{cases} \quad (1)$$

Where, S_k is the amplitude value segments for a movie. Histogram is calculated using these values. It is observed that thick portion in the histogram, which represents the video song sequences whereas other portions may have dialogs or actions scenes. Therefore, it is indispensable to remove unwanted parts. To curb it, threshold is computed, which is explained in following section

C. *Threshold Calcuation*

There are various statistical parameters are experimented to find out better threshold value. Root means square is worked well as compared to other parameters. It is calculated using equation 2.

$$P_{rms} = \sqrt{\frac{\sum_{k=1}^{n} s_k^2}{n}} \quad (2)$$

P_{rms} is the root mean square value and n is the total number of segments in a movie. This threshold is used to find ranges of video songs sequences. Extracted indexes may contain song sequence or action sequence or background music. It is observed that length of Indian Hindi video song sequence is around 3-4 minutes whereas action sequences and background music length is 1-2 minutes. Thus, ranges, which are less than 90 seconds, are removed and remaining sequences are considered as final song sequences. Eventually, these indexes contain only song sequences. It is useful for content based indexing and retrieval of song sequences in large movies and video song sequence databases.

III. EXPERIMENTAL RESULTS

To check diversity of our proposed approach , experimental results are carried out using various types of movies , which are shown in table 1. It shows that movie name, Genre, Language, Length of a movie and number of video song sequences. Length of movie is represented in hours , minutes and seconds.

Table -1 List of movies

Sr. No.	Movie Name	Genre	Language	Length of Movie(hh:mm:ss)	Number of Songs
1	Aitrazz	Drama, Thriller	Hindi	02:40	6
2	Ajanbee	Drama, Crime fiction	Hindi	02:45	7
3	Bazzigar	Drama, Thriller	Hindi	03:02	7
4	Chnennai Express	Romantic, Comedy	Hindi	02:21	5
5	Humko Deewana Kar Gaye	Romantic, Drama	Hindi	02:35	8
6	Jannat	Romantic, Crime Fiction	Hindi	02:14	6
7	Life Is Beautiful	Drama Romantic	Telugu	02:49:41	5
8	Koncham Ishtam Koncham Kashtam	Comedy Family Drama	Telugu	02:45:21	9
9	BeyYaar	Drama Comedy	Gujarati	02:30:36	5
10	Kevi Rite Jaish	Comedy Family	Gujarati	02:02:27	6

Video song sequences are extracted from our proposed approach ,which is shown in table 2. There are 5 video song sequences are extracted from moive 1. Start index of first video song sequence is 163rd segment and end index is 193. Similarly, Fifth video song sequence is started from 812nd and is lasted upto 833. Total number of video song sequnces are 6 in movie 1. Five song sequences are identifed correctly and one is misssed out.

Table -2 Video song sequences of movie 1

Song Sequence No.	Song_start_index	Song_end_index
1	163	193
2	262	289
3	347	374
4	525	552
5	812	833

Video song sequences of movie 2 are represented in table 3. Six video song sequences are identified from it. First video song sequence is started from 83rd segment and last segment number is 110. Total number of video song sequences are 7 in movie 2. Correctly extracted song sequences are 6 and 1 is missed out.

Table -3 Video song sequences of movie 2

Song Sequence No.	Song_start_index	Song_end_index
1	83	110
2	155	172
3	235	257
4	477	500
5	613	639
6	792	841

Experimental results are depicted in table 4. N_c represents number of video songs sequences, which are identified correctly using our proposed method, N_f depicts falsely identified video song sequences and missed video song sequences are represented in N_m column. Precision and recall is computed to find out average accuracy of our proposed approach. 58 number of video song sequences are detected correctly from our dataset whereas 11 song sequences are detected incorrectly. 7 video song sequences are missed out. Eventually, 87% precision and 98% recall are obtained in our approach.

Table -4 Simulation results of our proposed approach

Movie No.	N_c	N_f	N_m	Precision	Recall
1	5	0	1	1	1
2	6	0	1	1	1
3	5	0	2	1	1
4	5	1	0	0.83	1
5	7	2	1	0.77	0.86
6	5	2	1	0.71	1
7	5	0	0	1	1
8	9	3	0	0.75	1
9	5	0	0	1	1
10	6	3	0	0.67	1
Total	58	11	7	0.87	0.98

IV. CONCLUSION

The proposed system is divided into different modules. Video song sequences are extracted from long Indian movie using audio based approach. Initially movie is divided into fixed length of segments. Amplitude values are plotted for different movie in our analysis to find out global threshold value. Root mean square value is chosen for threshold after experiment with other statistical parameters. The system has been tested on 10 popular Indian movies of diverse genres. The selected movies contain a total of 58 songs out of which 47 were successfully extracted. 87 % accuracy has been achieved.

REFERENCES

[1] Wang, Y., Ostermann, J. and Zhang, Y.Q., 2002. Video processing and communications (Vol. 5). Upper Saddle River: Prentice Hall..
[2] Arman, F., Depommier, R., Hsu, A. and Chiu, M.Y., 1994, October. Content-based browsing of video sequences. In Proceedings of the second ACM international conference on Multimedia (pp. 97-103). ACM.
[3] Asghar, M.N., Hussain, F. and Manton, R., 2014. Video indexing: a survey. framework, 3(01).
[4] Patel, B.V. and Meshram, B.B., 2007, December. Retrieving and Summarizing Images from PDF Documents. In International Conference on Soft computing and Intelligent Systems (ICSCSI-07), Jabalpur, India (pp. 27-29).
[5] Castrillón, M., Déniz, O., Guerra, C. and Hernández, M., 2007. ENCARA2: Real-time detection of multiple faces at different resolutions in video streams. Journal of Visual Communication and Image Representation, 18(2), pp.130-140.
[6] Lienhart, R., Kuranov, A. and Pisarevsky, V., 2003. Empirical analysis of detection cascades of boosted classifiers for rapid object detection. Pattern Recognition, pp.297-304.
[7] Essays, UK. (2013). Limitations of Text Based Image Retrieval Psychology Essay. Retrieved from http://www.ukessays.com/essays/psychology/limitations-of-text-based-image-retrieval-psychology-essay.php?cref=1 (21 January 2016)
[8] De Bruyn, P., Bain, K., Allardice, D. and Joshi, S., 2010. Frommer's India (Vol. 761). John Wiley & Sons.C.S. Lu, H.Y.M Liao, "Multipurpose watermarking for image authentication and protection," *IEEE Transaction on Image Processing*, vol. 10, pp. 1579-1592, Oct. 2001.
[9] The Beat and Tempo. http://musictheoryblog.blogspot.in/2007/01/beat-and-tempo.html (11 July, 2015).
[10] Subashini, K., Palanivel, S. and Ramaligam, V., 2012, July. Audio-video based segmentation and classification using SVM. In Computing Communication & Networking Technologies (ICCCNT), 2012 Third International Conference on (pp. 1-6). IEEE.

