International Journal of Recent Technology and Engineering (IJRTE)
ISSN: 2277-3878 (Online), Volume-8 Issue-4, November 2019

Generating Highlights of Cricket Video using Commentators and Spectators Voice

Mohammed Inayathulla, C Karthikeyan

Abstract*: Videos are one of the important and richest sources of data on internet. In this growing world of digital technology video summarization will be handy in analysing the video data. Recently Natural Language Processing has attracted more researchers to work to meet the current emerging challenges. Among the various issues, video summarization got more focus and in this regard, many applications and works have been evolved. Video Summarization is the process of creating a small video describing the actual video within short duration(s). The paper focuses on generating highlights of a cricket video by analysing the voice of commentator and spectators. The experimental results have shown good performance when compared with human generated summary.

Keywords* : Classification, Deep Learning, Excitement, Video Summarization

I. INTRODUCTION

Due to the rapid advances in technology and growing use of social network platforms huge amount of various forms of data is generated. Amongst them videos are one of the crucial source of data. Some thousands of videos are daily uploaded by the users on various platforms. Hence video summarization will be handy as it significantly reduces the user's time and effort. Video Summarization aims at creating synopsis of the video without disturbing the actual semantics of the video. Video summarization is a challenging task for researchers because different users have different preferences for summary. Hence generating the summary which satisfies the preferences of different users is a complex task. Video summarization is broadly of two types static summarization and dynamic summarization. Static video summarization creates a storyboard which is the collection of random images which are of more importance. Dynamic video summarization creates a short video from the original one. Video summarization can be both manual and automated. Manual video summarization is a technique where the video is inspected to find the interesting and important frames which form the summary of the video. Automated video summarization involves training the machine to understand the semantics of the video so that machine generates the final summary of the video. Complexity of video summarization lies in understanding the semantic relationship between the frames.

Video summarization can be both supervised and unsupervised task. In supervised learning initially the machine will be trained with labelled data and depending upon the learning accuracy machine will generate the synopsis of the video. The generic approach of video summarization is described in figure 1. In unsupervised learning the machine will design criteria to select the important frames which form the summary of the target video. Manually designed criteria can be used in both supervised and unsupervised techniques for summary generation. Few of the researchers have focused on generating abstractive textual summary for the given video.

II. LITERATURE SURVEY

In this paper [1] the authors proposed a divide and conquer approach for video summarization. All the video frames were assigned with score of importance and subset of frames which gives the highest total score was selected. To get the optimal score for the entire video three stage deep neural network was proposed (2D CNN,1d FCN,LSTM). To deal with the subjectivity of summarization harmonic mean among different summaries was calculated. To evaluate the summary results were compared with the human created summaries. In this paper [2] authors proposed a DSSE (deep side semantic Embedding) model to generate summary. Side information available with video like title, query, comment, description was effectively used to find frames which form the summary. In this paper [3] the authors focused on users major or special interest for generating the summarization. The authors used Deep Ranking Model to evaluate the relationship between summarized and non summarized frames of the video.

In this paper [4] the authors developed a deep neural network for generating abstractive text summarization of the input video sequences. Initially the description of the input video was generated using CNN and from the description abstractive summary was generated. A Maximal Biclique Finding (MBF) algorithm [5] is devised in order to resolve the issues that raised due to the dearth of co-occurrence of patterns when irrelevant shots in videos are considered. The objective of this algorithm is to optimize the patterns that are located lightly and removing the less co-occurring patterns yet irrespective of their prevailing nature in one video. In [6][13][14] authors used

Manuscript published on November 30, 2019.
* Correspondence Author
Mohammed Inayathulla*, Research Scholar, Department of Computer Science and Engineering, Koneru Lakshmaiah Education Foundation, Vaddeswaram, AP, India. and Assistant Professor, Department of Computer Science & Engineering, Malla Reddy Engineering College (Autonomous), Hyderabad, Telangana, India. Inayathulla512@gmail.com.
C Karthikeyan, Associate Professor, Department of Computer Science and Engineering, Koneru Lakshmaiah Education Foundation, Vaddeswaram, AP, India. ckarthik2k@gmail.com.

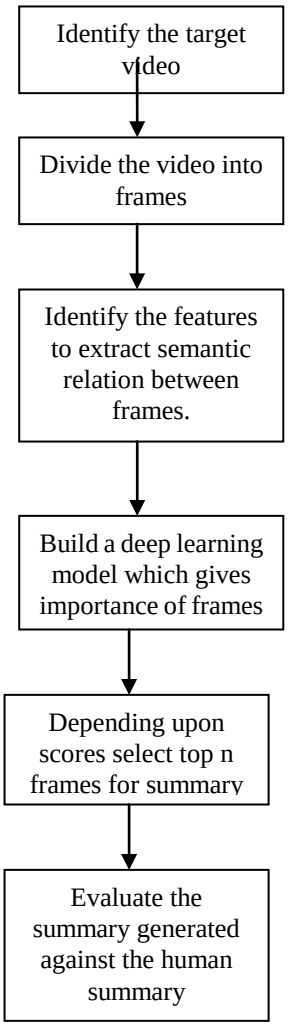

Fig. 1.Supervised approach of Video Summarization

Hidden Markov Model and empirical mode decomposition for generating summary. Initially the key frame was extracted then classification was applied to include it in highlights or not and then finally highlights was generated. In [7][15] authors used oriented fast rotated brief approach to find excited clips and optical charater and replay recognitions also were used. The authors proposed context aware video summarization [8] which uses sparse coding by learning from dictionary of video features and spatiotemporal feature correlation graph.

The authors proposed a hierarchical Recurrent neural network model for generating video summary [9]. The RNN was designed in such a way that the first layer was used to encode short video subshots which were taken from the original video and the second layer was used to calculate the confidence of each sub shot to form the summary. In [10][11][12] a novel approach for exploring surveillance video data. Initially object movement was recorded i.e. by extracting single shot image which is a combination of multiple frames. Action parsing driven video summarization model was proposed which was reinforcement learning based. For video summarization initially shot boundary detection was conducted then screen shot clustering was implemented using minimum spanning tree.

III. PROPOSED METHODOLOGY

To generate the highlights of cricket video voice of commentator and cheers of spectators are considered as key aspects. The general perception in this context is that when there is increase in commentator's voice or spectators cheers only when something important has happened, like a wicket or a batsman scoring four or six. The proposed approach is as described in figure 2.

After breaking the audio into chunks short term energy is computed and a threshold value for the energy is identified for classification of audio clips as excited or not. The threshold value can be experimented for improving performance. All the excited clips are merged together to generate highlights of the target video. To generate highlights parts of the video are to be extracted within a particular time interval only as the voice in the video is high only when the player performs an action.

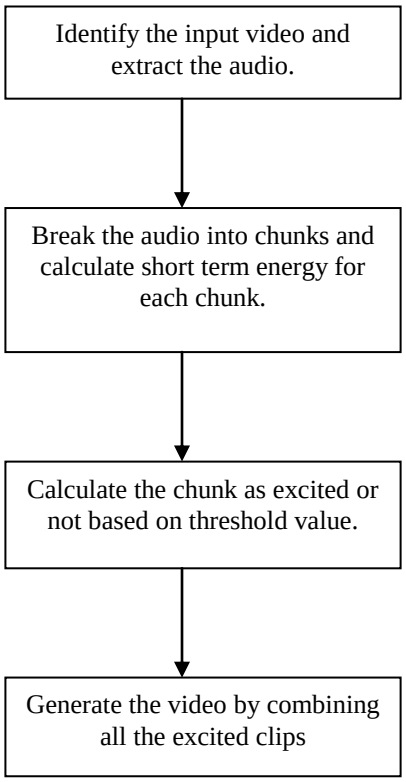

Fig. 2. Step by Step Process of Generating Highlights of a Video by Using its Audio.

IV. RESULTS

The proposed approach is implemented on cricket videos available on YouTube. A video of 23 minutes whose highlights in the form of excited clips was generated is shown in figure 3. Similarly we have implemented the work on 10 similar cricket videos which are downloaded from YouTube and the results are evaluated against the human generated summary. Selection of threshold value for selecting the excitement clip has severe impact on the performance of the proposed approach.

Fig. 3.A Snapshot of all the excited clips of Video.

V. CONCLUSIONS

The proposed approach which is based on voice of commentator and cheers of spectators has shown promising results in generating the highlights of a video. The proposed approach is limited to small size videos. In order to deal with videos of larger size or to generate the highlights of complete match deep learning techniques can be implemented which will improve the overall performance of the proposed approach.

REFERENCES

1. Siyu Huang, Xili, Zhongfei Zhang, Fei Wu, Junwei Han, "User-Ranking Video Summarization With Multi-Stage Spatio–Temporal Representation", IEEE Transactions on Image Processing, vol. 28, no. 6, pp. 2654-2664, June 2019.
2. Yitian Yuan, Tao Mei, Peng Cui, Wenwu Zhu, "Video Summarization by Learning Deep Side Semantic Embedding", IEEE Transactions on Circuits and Systems for Video Technology, vol. 29, no. 1, pp. 226-237, Jan 2019.
3. T. Yao, T. Mei, and Y. Rui, "Highlight detection with pairwise deep ranking for first-person video summarization," in Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 982–990, 2016, June 2016.
4. Aniqa Dilawari, Muhammad Usman Ghani Khan, "ASoVS: Abstractive Summarization of Video Sequences", IEEE Access, vol. 7, pp. 29253-29263, March 2019.
5. W.-S. Chu, Y. Song, and A. Jaimes, "Video co-summarization: Video summarization by visual co-occurrence," in Proc. IEEE Conf. CVPR, pp. 3584-3592, 2015.
6. Kamesh Namuduri, "Automatic extraction of highlights from a cricket video using MPEG-7 descriptors", Communication Systems and Networks Proc IEEE Conf , vol. 1, no. 1, pp. 1–3, 2009.
7. Daniel Ringis and Akash Pooransingh, "Automated Highlight Generation from Cricket Broadcasts Using ORB", IEEE Conf pp 58-63, 2015.
8. S. Zhang, Y. Zhu, and A. K. Roy-Chowdhury, "Context-aware surveillance video summarization." IEEE Trans. Image Process., vol. 25, no. 11, pp. 5469-5478, 2016.
9. B. Zhao, X. Li, and X. Lu, "Hierarchical recurrent neural network for video summarization," in Proc. ACM Multimedia, pp. 863-871, 2017.
10. Amir H. Meghdadi, Member, IEEE, and Pourang Irani; "Interactive Exploration of Surveillance Video through Action Shot Summarization and Trajectory Visualization"; IEEE Transactions On Visualization And Computer Graphics, vol. 19, no. 12, pp. December 2013.
11. Jie Lei ; Qiao Luan ; Xinhui Song ; Xiao Liu ; Dapeng Tao ; Mingli Song, "Action Parsing-Driven Video Summarization Based on Reinforcement Learning", IEEE Transactions on Circuits and Systems for Video Technology, vol. 29, no. 7, pp. 2126-2137, 2019.
12. Yihong Gong; "Summarizing Audiovisual Contents of a Video Program"; EURASIP Journal on Applied Signal Processing, vol. 2, pp. 160-169, 2003.
13. Gwang-Gook Lee, Hyeong-ki Kim, Whoi-Yul Kim, "Highlight Generation For Basketball Video Using Probabilistic Excitement", IEEE Intl Conference on Multimedia,2009.
14. Muhammad Rafiqul Islam, Manoranjan Paul, Michael Antolovich, Ashad Kabir," Sports Highlights Generation Using Decomposed Audio Information", IEEE International Conference on Multimedia & Expo Workshops,pp.579-584,2019.
15. Aman Bhalla ; Arpit Ahuja ; Pradeep Pant ; Ankush Mittal ," A Multimodal Approach for Automatic Cricket Video Summarization", 6th International Conference on Signal Processing and Integrated Networks (SPIN),2019.

AUTHORS PROFILE

Mohammed Inayathulla having more than 5 years of teaching experience in Computer Science and Engineering in India. He has been awarded with Masters Degree in Computer Science and Engineering from Ph.D. in Computer Science & Engineering from Jawaharlal Nehru Technological University (JNTU) Anantapur. He has published 5 papers in various International / National Journals, Conferences. His area of interest include Data Mining, Machine Learning. He is having professional membership in CSI and ISTE associations.

Dr. C. Karthikeyan having more than 18 years of teaching experience in Computer Science Engineering background in India and Abroad. He has been awarded Ph.D. in Computer Science & Engineering from Jawaharlal Nehru Technological University (JNTU) Hyderabad. He has published more than 30 papers in various International / National Journals, Conferences and book chapters. He organized various Seminars, workshops, symposium and conferences in India and abroad. He is having Professional membership in IEEE, IACSIT, ACEEE, ISTE, SDIWC and IAENG Associations.