ISSN: 0974-5823 Vol. 7 No. 2 February, 2022

International Journal of Mechanical Engineering

Imprinting Online Video Timeline Based on a User Keyword Concernment

[1]K.Siva rama Krishna, [2]B.V.Satish Babu, [3]Dr. A.Srinivas Rao, [4]K.Srinivas Rao

[1,2,3,4]Andhra Loyola Institute of Engineering and Technology,Vijayawada

Abstract - Online video content generation and the habit of watching online videos have been increased exponentially over the last few decades. In the scenario of lengthier videos, the user has to spend a lot of time over unnecessary content. Until now, we have not had any feature in online video platforms to identify the user portion of interest in the video timeline. To overcome the issue faced by online users, we put forward our innovative model for imprinting online videos based on a user keyword of interest. The user-concerned keyword is used to mark hotspots on the online video timeline. In this model, we are going to detach the audio line from the video. After separation, we are going to find the number of keyword occurrences along with their timestamps in the audio timeline. These timestamps are used to imprint hotspot points within the video. This feature enables and encourages online users to watch many videos in less time.

Index Terms - hotspot, cloud, transcribe, video, audio.

INTRODUCTION

The rapid development of digital video and editing technologies in today's world has led to an exponential increase in video data, demanding the use of effective and advanced techniques because multimedia repositories have made browsing, delivery, and analysis easier.

Users may prefer to watch the video summary, which is merely an abstract of the original video, rather than the entire video, which includes additional information about the occurrence of various incidents in the video. Therefore, summarizing a video appears to be an important procedure.

In the case of longer videos, the user must waste a lot of time watching irrelevant information [1]. There hasn't been a tool in online video platforms to detect the user's portion of interest in the video timeline until now.

We proposed a new methodology for imprinting online videos with hotspot markers based on a user term of interest. This method addresses the problem encountered by online users. In this paper, we have proposed one of the video summarization methods based on user keyword concernment.

RELATED WORK

Our main goal is to use the user's phrase of interest to imprint or designate a hotspot on the video. The audio from the video that piqued the user's interest is preserved in the AWS S3 bucket. We used Amazon Transcribe, which is based on deep learning, to convert the audio from video to text. Finally, using the Bootstrap video player, the hotspot is imprinted in the video timeline.

The rest of this paper is laid out as follows: In Section III, the Amazon transcribe is discussed. In Section IV, the literature is discussed. In Section V, proposed architecture is discussed. The outcomes are the focus of the VI section. The VII section is all about comparisons and analyses. In Section VIII, limitations and future work is explained.

AMAZON TRANSCRIBE

Amazon Transcribe is automatic speech recognition (ASR) service that allows developers to quickly convert speech to text. You can easily generate a transcript of an audio or video file that you uploaded to Amazon S3 by using the Amazon Transcribe plug-in.

It adds punctuation and formatting automatically, so the output is as good as any manual transcription. For more accurate transcriptions, you can set it up with custom vocabulary. You can even set it up to exclude certain words from transcripts. Many languages are offered by Amazon Transcribe.

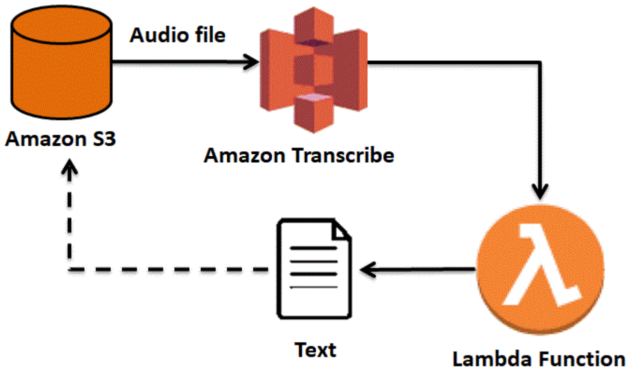

Figure.1 Amazon's transcriber's operation

A Lambda function is used to convert speech to text while utilizing Amazon Transcribe. AWS Lambda is a compute service that lets you run code without having to set up or manage servers. You can run code for almost any type of application or backend service with Lambda, and you don't have to worry about administration.

Editors can then use a text search to retrieve the file and include it in their work. The Fig.1 illustrates how Amazon transcribe operation.

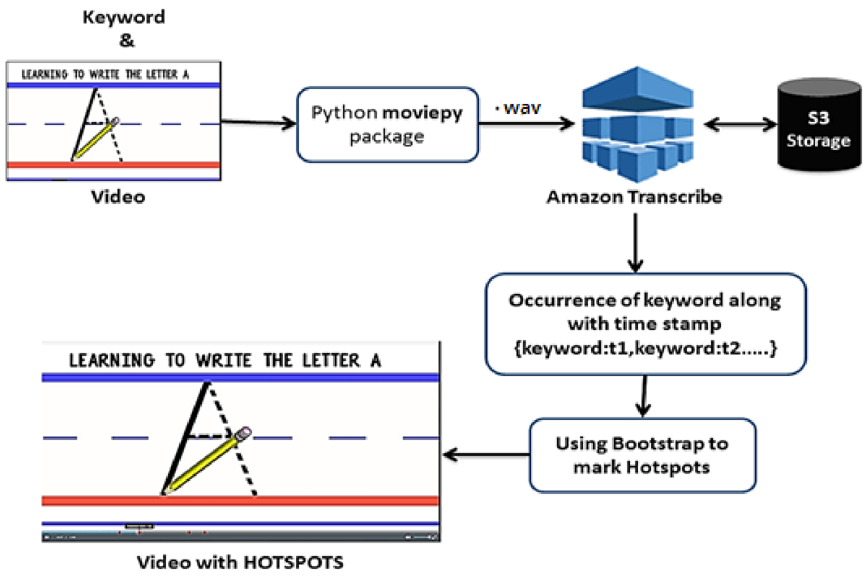

Figure.2 imprinting the hotspots on video timeline

LITERATURE

In this section, we've compiled a collection of different video summarizing and labeling research studies.

1) Bin Wu et.al. of [2], developed an automatic and unsupervised methodology for creating time- synchronized video tags from crowd sourcing comments information.

2) Yikun Xian et.al of [3], with the use of the video sentimental feature of time-sync comments, developed an automatic approach for extracting highlighting snap shots in video.

3) Yao T et.al of [4], proposed a cost-effective method of video labeling based on click-through information that can be considered as the imprints of client seeking behavior.

4) W. Yang et.al of [5], primarily concerned with obtaining video tags from time sync comments. A time-sync video comment is an interactive comment form in which users can make comments that are synced with the playback time of a video.

5) J. Li et.al of [6] suggested a model for detecting events in videos based on Time-Sync comments left by online users. Three aspects of Time-Sync comments are extracted initially in the model. The significance of user activity in time series is then examined to determine which video shots people are most interested in.

6) L. Chen et.al of [7], developed a method in which the user can choose a character in a video and deploy his or her emotion in the video timeframe along with a color bar.

The main idea of this work is based on the research journals mentioned above. All of these research strategies employ complex video processing and mining classifiers to focus on video tagging and summarization and given priority to user interest.

None of these studies focused on reducing the amount of time users spend watching videos. Our proposed solution focuses mostly on tagging as well as on reducing user spending time on watching videos.

PROPOSED ARCHITECTURE

Cloud services are now playing a pivotal role in addressing

many societal problems and various technical problems [9]. The method which has been proposed in Fig.2., is based upon the following cloud technologies like “AWS Transcribe cloud service”, “AWS S3 storage”. Additionally we have used two python packages “moviepy”, “boto3” python along with bootstrap video player.

The following steps are be used to explain the proposed method.

Step1: A keyword will be chosen based on the user's concern and the target video will be used as an input.

Step2: The video's audio line is retrieved using the Python package "moviepy.". Package“Boto3” is used to connect AWS S3 storage bucket from python script.

Step3: This audio line is sent to “Amazon transcribe,” which converts it to a lexicon of words (placed in S3 storage)

Step4: Locate keyword occurrences and get time stamps from a word dictionary.

Step5: Using the bootstrap video player, create HOTSPOT markings at the appropriate timestamps in the video player.

RESULTS

Our preferred framework is tested on a variety of videos of different lengths. In each run, the user will be asked for two inputs, 1) keyword of concern 2) Target video, as shown in Fig. 3

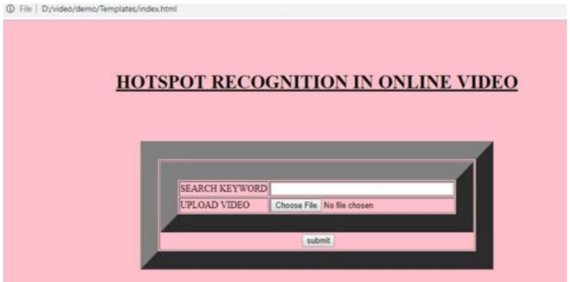

Figure 3 Two inputs from the user

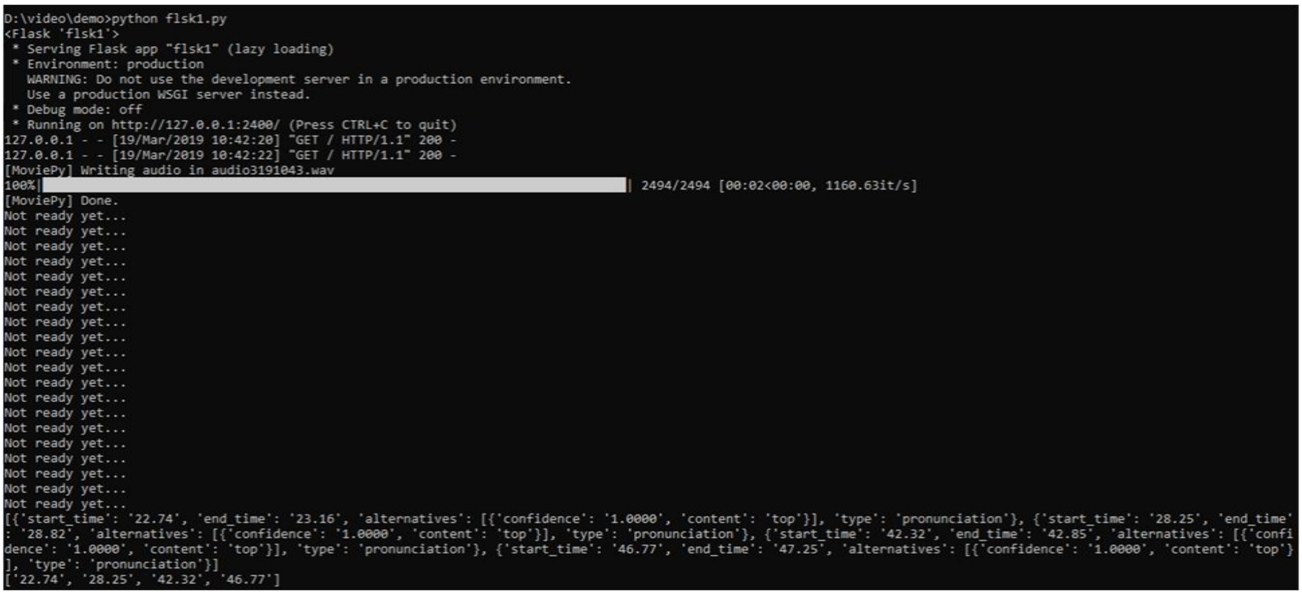

Figure.4 “top” Keyword occurrences along with timestamp

After receiving two inputs from the user, the “moviepy” package is used to generate an audio file in .wav format and that audio file is uploaded to the S3 bucket, as shown in Fig.5

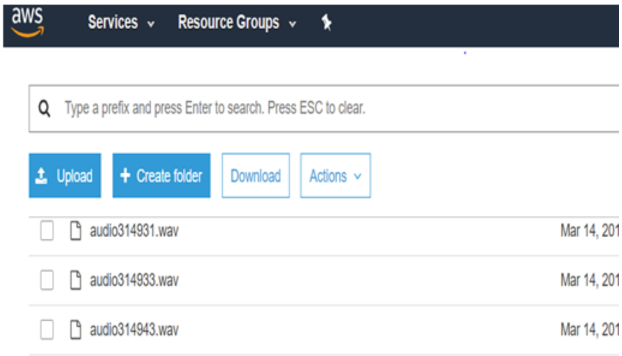

Figure.5 S3 Storage with retrieved audio files

As illustrated in Fig.4, from the .wav audio file, the occurrences of the example keyword “top” are listed out along with their timestamps in the dictionary format. The Keyword “top” occurrences along with timestamps and confidence values are actually stored and retrieved from AWS S3 storage as shown in Fig.6

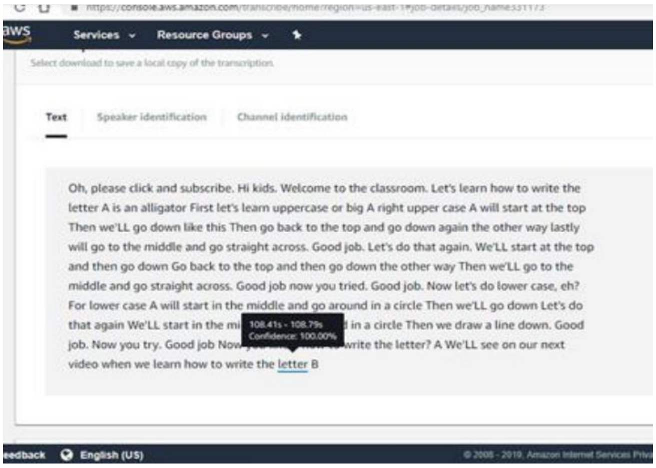

Figure. 6 Data in AWS S3 storage bucket

Every word in Amazon Transcribe output has a confidence score. The confidence result represents the accuracy of a word transcribed. Each word gets a confidence score ranging from 0 to 1.

The timeline of the video is imprinted with red color marker points at all the 4 occurrences of the keyword "top". These marker points are made based on the timestamps acquired from the deep learning-based Amazon transcribe service. The final result is depicted in Fig.7.

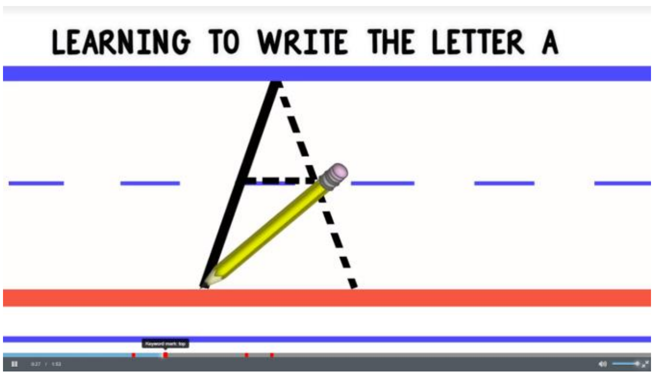

Figure. 7 imprinting the marker points on user audio

The same procedure has been used to test a variety of example videos, such as Fig8. Markers placed on the video timeline are compared to text acquired from Amazon transcribe in every case.

Figure. 8 imprinting the marker points on sample video

Table-I Analysis information about video samples

Video number	Video length	Time taken to mark	Number of markers	Number of false positives
Video sample_1	08:33	01:15	15	2
Video sample_2	01:53	00:45	4	1
Video sample_3	04.50	00:55	2	0
Video sample_4	03.23	00:25	6	0
Video sample_5	06:00	01:02	6	1
Video sample_6	03:45	00:44	7	1
Video sample_7	07:15	01:25	9	3
Video sample_8	01:53	00:42	5	0
Video sample_9	05:20	01:57	10	2
Video sample_10	02:30	00:10	4	0

DISCUSSION

Different video samples were used to test our suggested technique. We have notable outcomes in every situation, with extremely few false positives. Table-I displays the results of ten video samples. The following analysis is based on the findings:

The number of markers is entirely dependent on the keyword type; video length has no bearing on the count of markers.

As the number of markers grows, the number of false positives grows slightly.

It was discovered through cross-verification that false positives are occurring as a result of enabling partial matching. The accompanying line chart (Fig.9) reflects these findings.

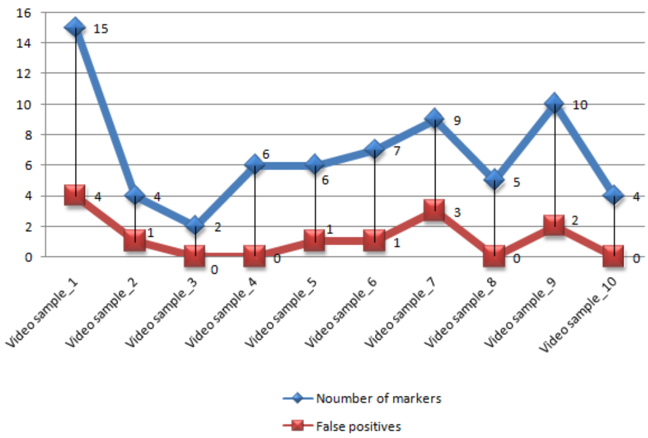

Figure.9 Mapping of false positives and markers

LIMITATIONS & FUTURE WORK

Currently, our proposed method is used to process videos that are uploaded to the application. We plan to use the same strategy for internet videos in the future. We also recommend that machine learning algorithms be used to reduce the number of false positives.

CONCLUSION

.In today's digital era, there is an exponential increase in video content because most of the users are interested in watching video content. To reduce the user time spent over watching unnecessary parts of the video, we have introduced the method of imprinting an online video timeline based on a user keyword concernment. Our method has been tested on different videos, and it has proven effective with fewer false positives

REFERENCES

[1] S. Tan, H. Tan, and C. Ngo. Topical summarization of web videos by visual-text time-dependent alignment. In MM, pages 1095C1098. ACM, 2010.

[2] Wu B, Zhong E, Tan B, et al. Crowd sourced time-sync video tagging using temporal and personalized topic modeling, Proceedings of the 20th ACM SIGKDD international conference on Knowledge discovery and data mining. ACM, 2014: 721-730.

[3] Xian Y, Li J, Zhang C, et al. Video Highlight Shot Extraction with Time-Sync Comment,Proceedings of the 7th International Workshop on Hot Topics in Planetscale Mobile Computing and online Social Networking. ACM, 2015: 31-36.

[4] Yao T, Mei T, Ngo C W, et al. Annotation for free: Video tagging by mining user search behavior,Proceedings of the 21st ACM international conference on Multimedia. ACM, 2013: 977-986.

[5] W. Yang, N. Ruan, W. Gao, K. Wang, W. Ran and W. Jia, "Crowdsourced time-sync video tagging using semantic association graph," 2017 IEEE International Conference on Multimedia and Expo (ICME), 2017, pp. 547-552, doi: 10.1109/ICME.2017.8019364.

[6] J. Li, Z. Liao, C. Zhang and J. Wang, "Event Detection on Online Videos Using Crowdsourced Time-Sync Comment," 2016 7th International Conference on Cloud Computing and Big Data (CCBD), 2016, pp. 52-57, doi: 10.1109/CCBD.2016.021.

[7] J. Chen, W. Wu, W. Hu and L. He, "TSCREC: Time-sync Comment Recommendation in Danmu-Enabled Videos," 2020 IEEE 32nd International Conference on Tools with Artificial Intelligence (ICTAI), 2020, pp. 67-72, doi: 10.1109/ICTAI50040.2020.00021.

[8] L. Chen, G. Chen, C. Xu, J. March and S. Benford, "EmoPlayer: A media player for video clips with affective annotations," in Interacting with Computers, vol. 20, no. 1, pp. 17-28, Jan. 2008, doi: 10.1016/j.intcom.2007.06.003.

[9] Babu, B.V.S., Babu K.S. (2021). The purview of blockchain appositeness in computing paradigms: A survey. Ingénierie des Systèmes d'Information, Vol. 26, No. 1, pp. 33-46. https://doi.org/10.18280/isi.260104..