Lecture Video Summarization

Durjoy Ghosh
PES1201802124
Computer Science Department PES University, RR campus BSK III, Bangalore, India
durjoy12091999@gmail.com

Vrushabh Chougale
PES1201801495
Computer Science Department PES University, RR campus BSK III, Bangalore, India
1008vrushabhchougale@gmail.com

Shreya Deepak
PES1201801099
Computer Science Department PES University, RR campus BSK III, Bangalore, India
shreyad1999@gmail.com

Surabhi Shivanand
PES1201800849
Computer Science Department PES University, RR campus BSK III, Bangalore, India
surabhi251000@gmail.com

***Abstract*— Universities, and educators upload a multitude of educational videos online. Lecture Videos are more prevalent in the present era citing Covid-19. These videos tend to be long and contain too much information to consume. A Lecture Video Summarization system takes an input lecture video and converts it into relevant slides. This system helps users see a summarized version of a longer video in the form of a presentation. They can even go through the relevant topics in an easy manner and hence make revisions quicker. This way the time and effort of the user are saved. This is done using shot detection, video splitting, and extracting keyphrases using the N-gram method. Later, everything is integrated together and the detected shots are used to make slides. The main aim of the project is to make life easier for students by letting them access the information in the lecture videos and brush up on topics just before exams.**

Keywords—summarization, keywords

I. Introduction

Video summarization is a process of distilling a raw video into a more compact form like video, audio, text, etc. without losing any information. This can be done by selecting informative frames of the video and capturing relevant information.

There are two methods for video summarization - unsupervised and supervised methods. In unsupervised video summarization, heuristic approaches are used to select the keyframes. The methods used in unsupervised video summarization are keyframe extraction and video skimming. Keyframe extraction is a process where a set of frames are extracted from a video. Video Skimming is a method that makes the summarized video more continuous and decreases the number of abrupt changes of frames and cuts. In supervised both raw video and its summarized video is given for learning. The disadvantage of using supervised learning in summarization are the shortage of the summarized video dataset. To summarize a video, one has to sit through the whole video and select keyframes from that video, which is a very time-consuming process. The main aim is to produce a summarized video by selecting important parts of the video which represent the whole video.

Some of the applications of video summarization are CCTV Surveillance Video Summary, Sports Video Highlights Generation, Lecture Video Summarization, and many more. The application that we are focusing on is lecture video summarization.

One of the most important aspects of e-learning is the digitization of scholarly articles which includes recording lecture videos. In recent times there has been a surge in the number of lecture videos being recorded by the universities with the pandemic being one of the driving forces. These

lecture videos enable the students to use the videos whenever they want to study a particular topic. Since the lecture videos are lengthy, students face difficulty going through the lecture videos. The videos might contain content that is not relevant to the topic. The users will have to go through the entire video to search for the required sub-topics which are time-consuming. Quick revisions during the exams are also not possible with the whole video in place. This project aims to summarize the lecture videos' content to save time for the user. This helps the students with their quick revision during exams, find particular subtopics of interest in the video, and save ample time.

II. Scope

We are developing a system for video summarization. This system can be utilized for summarizing video lectures. Since lecture videos are pretty lengthy, it is very difficult for the students to surf through the videos to search for a particular topic. Also, quick revision during the exams is also not possible with the whole video in place as it contains many things which are not so important. The summarization technique can help the students to find a particular topic in the whole video very easily. Since we are summarizing the whole video into a slide show with the most important snapshots of the lecture helps the students with a quick revision. This system can be given any lecture video, which it extracts the audio from and converts it to text, and then summarizes the text. The timeline of the summarized text will be matched with the important snapshots taken.

III. Literature Survey

[1] Lecture Video Summarization Using Subtitles

Overview:
In this paper, first, the content of the elements is extracted from the video and these frames are mixed together to produce the summarized video. Text summarization is done by audio extraction from the lecture video and subtitles are added. Then, the summary of the video is obtained. Here, four modules are used to obtain the summarized video - conversion module, preprocessing module, text summarization module and video summarization module.

The evaluation of the summarized video is done by using ROGUE(Recall-Oriented Under- 257 study for Gisting Evaluation) toolkit.
Some improvements can be made in the proposed method to increase its efficiency when punctuations are used. This can be done with the help of binary classification of the pauses between the utterances of the user and the parts of the video that can be removed. This results in a smooth transition of video frames and a meaningful summarized video can be obtained.

[2] Whiteboard Video Summarization Via Spatio-Temporal Conflict Minimization

Overview:
In this paper, the lecture videos are summarized by using lesser frames which are based on reducing conflicts between regions having content. A spatio-temporal index is used to navigate through lecture videos. The efficiency of the method is evaluated by using four baselines - binary, whiteboard segmentation, ground truth-based whiteboard segmentation, and reconstructed.

It is observed that the binary frames obtained the highest recall but had a lesser precision because of all the non-content connected components. The limitation in the paper is that the proposed method is not efficient when the lectures are recorded using multiple cameras.

[3]Automated Summarization of Lecture Videos

Overview:
In this paper, they have tried to design a tool that tackles problems such as storage archiving large

numbers of videos for a prolonged time, navigating among the different sub-topics taught in class, and helping the students by summarizing the videos so as to give the essential information without watching the whole lecture.

Hence this tool performs two main tasks that are splitting the lecture videos and then summarizing them. They found two limitations in this tool: one is topic overlapping and the other is that some of the important points in the original video are cut out in the summary video.

[4]Auto Summarization of Audio-Video Presentations

Overview:
In this paper, They discuss the different sources that can be used to obtain automated summaries. They are namely: video channel, audio channel, speaker's actions, and end-users actions. But they use only three out of the four sources: the audio channel, speaker's actions, and the end-users actions.

They used three algorithms to summarize the videos, the first algorithm uses information in slide transition information only, the second algorithm tries to detect emphasized regions in a speech by pitch activity analysis and the third algorithm (SPU) uses all three sources of information i.e., the slide transitions, pitch activity, and user-access patterns.

They then decided to test this with human subjects. Based on the properties of clarity, conciseness, and coherence, the computer-based methods were rated particularly poor on coherence. The subjects complained that the summaries "jumped topics" and were "frustrating because it would cut out or start in the middle [of a topic]," and so forth.

They concluded that they didn't find any major differences between users' preferences for the three computer-based methods that lead them to believe that the simpler methods (slide-based and pitch-based) may be preferable for now.

IV. METHODOLOGY

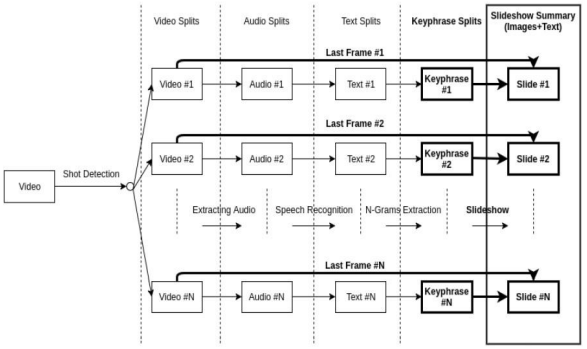
We split the video in a manner so as to separate the sub-topics taught in the lecture. Capture the moment in the video where the professor moves on to the next slide, it can be said that the professor has moved to the next sub-topic. It is this change in slide that is to be detected. Hence, split the video wherever a shot transition takes place. An additional benefit of detecting shots is that camera focus change from the professor (and the teaching board) to the slides can also be detected. This helps to capture content written on the black board as well.
During the lecture, professors generally design their slides in such a way that each point in the slides appears one by one as they speak. From each video split we extract the last image from the split which gives the summary of that particular split.

It is necessary to know what the professor is saying during the video lecture. Hence Audio is extracted from the video splits.

It is important to extract text from each audio splits to identify the key phrases used for each subtopic. Google speech API has been used for speech to text conversion.

Key Phrases have to be extracted for each text split as each text split can contain a different sub topic. To do so, convert the json file to raw text format. Remove punctuations and special characters etc. Remove stop words. Generate all 1-Grams and 2-Grams and record the frequency of each of them.

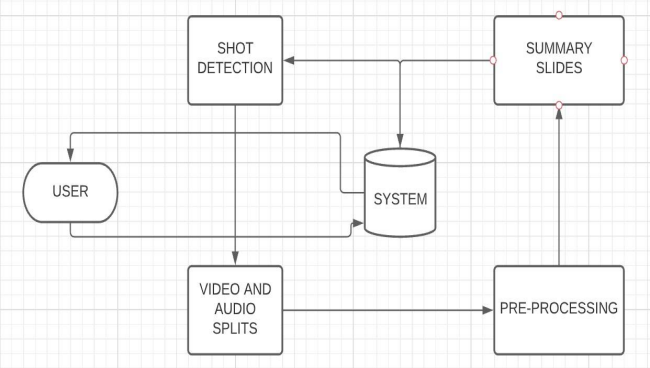
The slides for the lecture will be an ordered sequence of images extracted from each of the video splits along with the key phrases extracted.



V. DATA

The data to this domain can be any lecture video that has been recorded with loud and clear audio and video. Since this is an unsupervised method there is no need to have a dataset to train. The method works well with lecture videos of all lengths. But the lecture videos are supposed to be in English since that is the only API that has been employed in this method. The tests have been performed on video lengths ranging from 5 mins to 55 mins.

VI. IMPLEMENTATION



The system has been implemented in 6 modules. The input of the sequential module will depend on the output of the previous one.

1) **Image Extraction:** During the lecture, professors generally design their slides in such a way that each point in the slides appears one by one as they speak. From each video split we extract the last image from the split which gives us the summary of that particular split.

2) **Video Splits Generation:** Based on the shots detected, there are 'n' number of video splits generated. The number of video splits generated will be dependent on the drastic changes in the frames. This is because a new split is created for new changes in the frame.

3) **Audio Splits Generation:** It is necessary to know what the professor is saying during the video lecture. Hence Audio is extracted from the video splits. This is done using useful APIs.

4) **Speech to Text Translation:** The translation in the user specified language is initially done in the text format and later it is converted into audio format and passed to the user.

5) **Keyword/Key Phrase Extraction:** Key Phrases have to be extracted for each text split as each text split can contain a different sub topic. To do so, convert the json file to raw text format. Remove punctuations and special characters etc. Remove stop words. Generate all 1-Grams and 2-Grams and record the frequency of each of them.

6) **Summary Slide Generation:** The slides for the lecture will be an ordered sequence of images extracted from each of the video splits along with the key phrases extracted.

VII. RESULTS & DISCUSSIONS

- We have a website that helps user to summarize their lecture videos in the form of slides
- We have also made sure to attach the clip associated with the slide so that if students want an explanation about the topic(s) on the slide they don't need to go through the entire lecture
- We also have a search bar that helps students do a topic based search so that they don't need to go through all the slides to find the topic they are searching for.

VIII. Conclusion

Some of the applications of video summarization are Video Database Management, Consumer Video Analysis, CCTV Surveillance Video Summary, Sports Video Highlights Generation, Lecture Video Summarization and many more. The application that we are focusing on is lecture video summarization.

Our conclusions:

- The accuracy of the components individually is good, however, there is a huge scope for improvement.
- Because the components are executed sequentially, the overall accuracy decreases as the errors in each component get carried forward and affect the components next in line.

IX. Future Work

- We need to increase the accuracy of each component as well as the overall accuracy.
- We need to develop methods to filter and cover up the errors getting carried forward from the previous component.
- We need to ensure that the users have a great experience, checking if we can smoothen any process that will help the users to have a better experience.

References

1. Ramamohan Kashyap Abhilash, Choudhary Anurag,Vaka Avinash and D.Uma, “Lecture Video Summarization Using Subtitles”, 2020 2nd EAI International Conference on Big Data Innovation for Sustainable Cognitive Computing, EAI/Springer Innovations in Communication and Computing. https://doi.org/10.1007/978-3-030-47560-4_7
2. K. Davila and R. Zanibbi, "Whiteboard Video Summarization via Spatio-Temporal Conflict Minimization," 2017 14th IAPR International Conference on Document Analysis and Recognition (ICDAR), Kyoto, Japan, 2017, pp. 355-362, doi: 10.1109/ICDAR.2017.66.
3. A. Vimalaksha, S. Vinay, A. Prekash and N. S. Kumar, "Automated Summarization of Lecture Videos," 2018 IEEE Tenth International Conference on Technology for Education (T4E), Chennai, India, 2018, pp. 126-129, doi: 10.1109/T4E.2018.00034
4. He, L., Sanocki, E., Gupta, A., and Grudin, J., “Auto-Summarization of audio-video presentations”,1999 In Proceedings of the ACM Multimedia Conference (ACMMM), Orlando, FL.