Video Transcription in to Enhanced Text Summarization

E. Kamalanaban1, J. Senthil Murugan, Hema Sureya S, Karthik Raj V N, Kaviyan A

Vel Tech High Tech Dr.Rangarajan Dr.Sakunthala Engineering College, Tamil Nadu, India

Corresponding Author Email: j.senthilmurugan@velhightech.com

Abstract: *Every day, a huge quantity of videos are made and shared online. It's very tough to spend time watching videos in our lifestyle. First, we have to discover the right movie for our problems. If we don't figure out the solution from the video, we become frustrated, which has an adverse influence on our health. We simply require pertinent information from the film, but occasionally, in an attempt to attract visitors' interest, they post deceptive thumbnails, descriptions, advertisements, and other content! Every year, the number of creators rises quickly, which has an immediate effect on the quantity of videos that need to be made. Instead of providing accurate information, the authors may provide deceptive information because of their want for views. The creators of the videos squander our time by endorsing various businesses, reiterating the term for channel subscription, sharing the video, and soon! A Chrome plug in called Youtube Transcript Summarizer provides a time-capsulated summary of the entire video. We included a button in the extension that shows asynopsis of the current video playing on the Chrome browser to facilitate user- friendly interaction.*

Keywords: *Text Summarizer, Hugging Face Transformers, Chrome Extension, Web API.*

1. INTRODUCTION

These days, YouTube videos are a great source of information, built might be difficult to learn from them because it takes a lot of time to absorb the necessary knowledge. Videos may have unnecessary and useless content that either cannot be skipped or cannot be avoided, forcing the viewer to watch it. By 2023, there will be about 2.6 billion YouTube subscribers, and that figure will grow daily. Every hour, on average, 20,000 new videos are uploaded to YouTube, with an average duration of 12 minutes. The number of videos that are shared is growing constantly, and users cannot waste their valuable time watching videos for no purpose. Sometimes watching a video takes longer than anticipated, which makes it very tough, and other times the user's efforts are in vain if the movie doesn't provide the accurate information they were looking for. Trying to locate movies with reliable information on the topic we are searching for can be annoying and time- consuming. For viewers who would prefer to get correct information without investing their time in watching the movie, this summarizer is useful. To help users find accurate information and solutions to their issues, the summarizer provides a text description of the video's transcript. Because they provide a written representation of the audio in a video, transcripts are a great source of information. Natural language processing (NLP) approaches that extract important phrases and sentences from the transcript can be used to produce transcript summarization. After that, the created summary can be utilized to give visitors a brief synopsis of the video. The chrome extension is available for download from the store, and in order to use it, we must click on the icon when the video is playing. The popup will appear in the background, and we will need to click the button containing the summary of the title. In just one minute, the extension will show the video's content in a popup window! In this way, users' time will be saved by the addon. The summarizer is Chrome add on that syncs with YouTube to enable users to view and extract a video's essential parts. The summary can be tailored to the user's preferences, offering different levels of summarization. A modest user interface (UI) is used to show the user with the key points from the summarizing process to get her with the associated time-stamps next to the video feed. This makes it possible for the viewer to quickly and easily go to the most crucial parts of the video and its main ideas.

2. LITERATURE REVIEW

With so much video content available on the site, researchers have been paying more attention to the use of YouTube transcript summarizers in recent years. A review of some of the earlier studies on YouTube transcript summarizers is given in this section. Cuneyt M. Taskiran, Aronon Amir, Dulce B. Ponceleon, and Edward J. Delph's paper, "Automated Video Summarization Using Speech Transcript," explains how compact representations of video data might facilitate effective video browsing. They provide an algorithmic way for summarizing a lengthy film. Their representations retain the content's key elements while giving the user pertinent information about the specific sequence that was examined. The paper "Video Summarization using NLP" by Sanjana R., SaiGagana V,Vedhavati K.R.,and Kiran K.N. suggests employing algorithms based on Natural Language Processing (NLP) to automatically summarize videos. Due to YouTube's growing popularity, we now have access to millions of videos, which has increased demand for effective summarizing algorithms that can accurately summarize a variety of videos without sacrificing any of its value. Their suggested system explains the YouTube video transcripts that are used to create the summary video. YouTube, Reddit, Instagram, and other repositories host millions of videos that are generated and shared. It is getting harder to find the time to view these videos, some of which are longer than others. Summarizing the transcripts of such movies can assist us in obtaining the relevant information from the video, even though watching videos may not always be worthwhile if we are unable to learn anything valuable from them. The methodology for summarizing videos on YouTube is useful for extracting transcripts and creating summarized versions of them. The model automatically generates a summary that includes all pertinent details from the original material and includes key phrases. While the extractive strategy removes sentences and phrases from the input, the abstractive approach creates new words from the original text, making the work more challenging. A deep learning model-based approach for summarizing YouTube video transcripts was presented byNallapatietal.in2018.Convolutionalandrecurrent neural networks were combined in the method to extract pertinent information from the transcript. Using a dataset of YouTube videos, the authors assessed their approach and produced competitive results when compared to existing summary techniques. In 2019, Nguyen et al. presented a hybrid method that blends machine learning and rule-based techniques to summarize YouTube video transcripts. Sentences from the transcript were extracted using a set of rules, and a machine learning model was trained using the extracted sentences to provide a summary. Using a dataset of TED Talks, the authors assessed their approach and produced results that were competitive when compared to existing summary techniques. A graph-based method for summarizing YouTube video transcripts was proposed by Zeng et al. in 2020. The technique extracted important terms from the transcript by combining the TF-IDF and Text Rank algorithms. Using a dataset of TED Talks, the authors assessed their approach and produced results that were competitive with previous summary techniques. Huang et al. (2021) presented a transformer-based language model method for summarizing transcripts of YouTube videos. To extract pertinent data from the transcript, the technique made use of a transformer model that had been trained beforehand. Comparing their approach to existing summarizing techniques, the authors found that theirs produced competitive results when tested on a dataset of instructional videos. The literature review concludes that different methods, such as deep learning models, hybrid approaches, graph-based approaches, and transformer-based language models, have been used to approach YouTube transcript summarizing. These techniques have Demonstrated competitive outcomes when compared to alternative summarizing techniques. They have been tested on various datasets, such as TED Talks, instructional videos, and generic YouTube videos.

3. PROPOSED SYSTEM

The spoken text or the natural language context, which are two of the most significant information sources in a video sequence, are not used in the majority of video summarizing techniques. Speech recognition software can be used to extract transcripts from audio files for events like speeches, seminars, and educational programs that do not yet have them. These transcripts can then be used by our summarizer. A tool called YouTube Transcript Summarizer uses the audio transcript of a video to automatically create a summary. The model will entail creating and optimizing several methods and algorithms for information extraction and natural language processing (NLP), as well as putting them into practice and testing them on a sizable dataset of YouTube transcripts. This model uses a variety of APIs, including the Python API for retrieving YouTube videos and the FLASK API for testing, as well as a variety of languages and frameworks, including HTML, CSS, and JavaScript, to construct the extension for the web browser. Text summarization uses analytical techniques from Natural Language Processing (NLP) that are based on information extraction techniques. Using artificial intelligence algorithms, this method thoroughly analyzes the meaning of the source text to generate a source representation tailored to a certain

application. Next, using this source representation, a summary representation is constructed along with the summary text. However, methods that rely on statistical processing to extract phrases for the summary sometimes result in summaries that lack consistency. Another problem with these strategies is that dangling anaphors, which are pronouns, demonstratives, and comparatives like "he,""this," and "more," can only be understood in relation to an antecedent clause that occurs before the sentence in which they appear. When using anaphoras, the reader might not understand the summary if the antecedent clause is missing. Better summaries are produced by NLP-based algorithms, although these systems often require a large and intricate knowledge base. Furthermore, these systems are usually restricted to a specific application domain and difficult to develop into new areas. The goal of the artificial intelligence and computer science fields of natural language processing, or NLP, is to make computers able to comprehend and react to spoken and written human language. This essentails using methods from statistical modeling, machine learning, deep learning, and computational linguistics to process text or audio data, including human language, and extract the sentiment and intended meaning. NLP encompasses a variety of tasks, including part of speech tagging (identifying the part of speech of a word or text based on context), speech recognition (converting voice data into text), word sense disambiguation (figuring out the correct meaning of a word with multiple meanings), sentiment analysis (extracting attitudes, emotions, sarcasm, and other subjective qualities from text), and natural language generation (converting structured information into human language).Natural language processing, or NLP, aims to give computers the ability to understand and use language similarly to humans. There are many potential uses for this, such as voice assistants, chatbots, customer service, and more.

4. METHODOLOGY

The main goal of this essayis to give concise, accurate, and clear summaries of YouTube videos o that readers won't have to waste their time. For each goal, this paper uses more well used Pythonlibraries. These are the Google Translate API for translation, the YouTube Transcipt API for transcript extraction, the Flask framework for backend connections, and BERT, a summary package that combines GPT and BART.
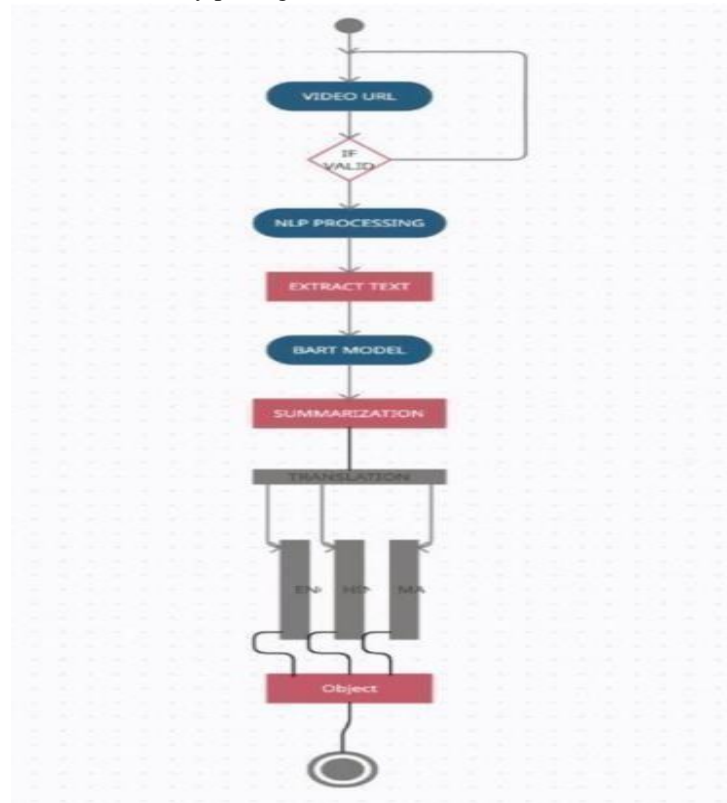

FIGURE 1.

Getting URL: A Uniform Resource Locator, or URL, is a pointer to an internet resource that tells you where it is on a computer network. It serves as a means of accessing information found on the internet. When a user types a YouTube URL into the search box, the system processes it in the background to collect the necessary data. Next,

as certain the legitimacy of the URL. In that case, the link process will be shortened by the backend. After that, the required YouTube video will open from the given URL. It will take care of text extraction from videos. We shall ascertain whether the URL is authentic when the user enters it. If the URL is invalid it will return the error message. In the event that the URL is invalid, an error notice will appear. If the URL is authentic, the correct URL will be supplied to the NLP model in the subsequent stage.

5. VIDEO TO TEXT EXTRACTION

After the required YouTube video has been obtained, the URL from the NLP model file will be sent to the utubeextract.py file. Next, we will use the youtube_transcript_api to get the transcript for the required YouTube video. There are three ways to generate transcripts: utilizing films without transcripts is the third approach, while the first is automatic Videos without transcripts won't be offered among our options. Once the transcript of a video has been obtained via an API, we parse the text in order to prepare ready for then extaction. Remove all commas, punctuation, and full stops because they are essential for indicating sentence boundaries. The "punctuator" Python library can be used to perform this task. The next step is to apply the text preprocessing algorithm for the transcript that was extracted. Eliminating stop words, exclamatory phrases, stemming, and punctuation for scanning purpose siscomparable to the previously discussed technique. The previously mentioned task is to summarize a long term in a brief manner. That overview of the video will provide important details. When summarizing text susing NLP, the Extractive Summarization is commonly employed. This extractive summarization only contains the most important words and sentences in its summary. The user or customer will receive immediate notification if the quantity specified changes as a result of the website's continuous monitoring of the data on the webpage they have provided. After the user enters the necessary data, our website will extract the necessary information from the product's webpage. Beautiful Soup Module is used by our website to scrape data. It is a Python package that is primarily utilized for online data scraping.
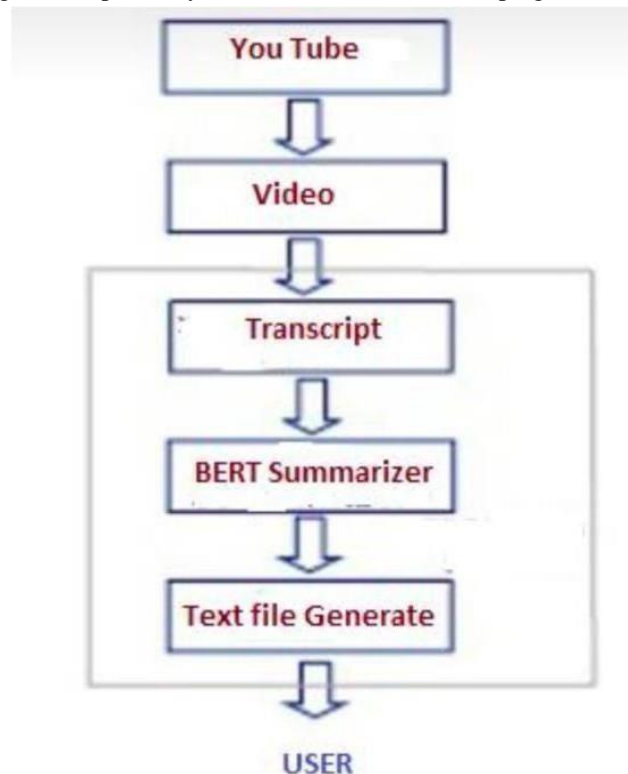

FIGURE 2.

2. NLP PROCESSING:

This process is for text summarization. In NLP, there are several ways to summarize text. We'll talk about text summarizing, which is the process of distilling along paragraph into a brief synopsis. We need more time to cover the content of a paragraph with numerous lines, but we only have time for the text's principal report. By removing

stuff that isn't needed, we can make the long paragraph shorter. The technique of breaking up a lengthy text into digestible paragraphs or phrases is known as NLP text summarization. The process outlined above will be used to get the important data.
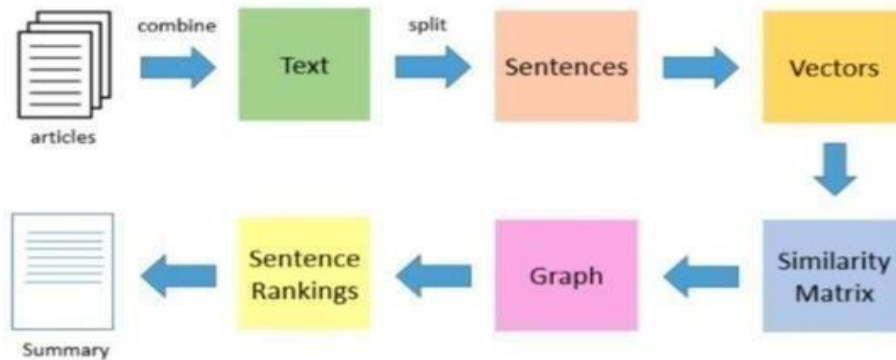

FIGURE 3.

3. BERT SUMMARIZER:

BERT stands for Bidirectional Encoder Representations from Transformers. This BERT offers a brand-new, state-of-the-art approach to finishing NLP jobs. BERT is one of the main algorithms used in models of natural language processing. The relevant information can be extracted while the essential information is retained by employing extractive text summarization. This time, it's more difficult to get this summary. In this effort, we are using improved embeddings provided by encoder models like BERT. An extractive summarizer built using two supervised approaches and the BERT sentence embeddings. Only the former consider embeddings and their derivatives. By examining the fundamental organization of the content, we can select pertinent, understandable sentences and provide a strong summary. The Unsupervised Text-Rank model is the conventional approach. New corpuses gain from the sequential information and well-known features in addition to these techniques. Actually, several magazines have employed this strategy. The baseline for the second approach is lead. For Rouge-1 and Rouge-L, the F1 measure is relevant when the supervised models perform better.
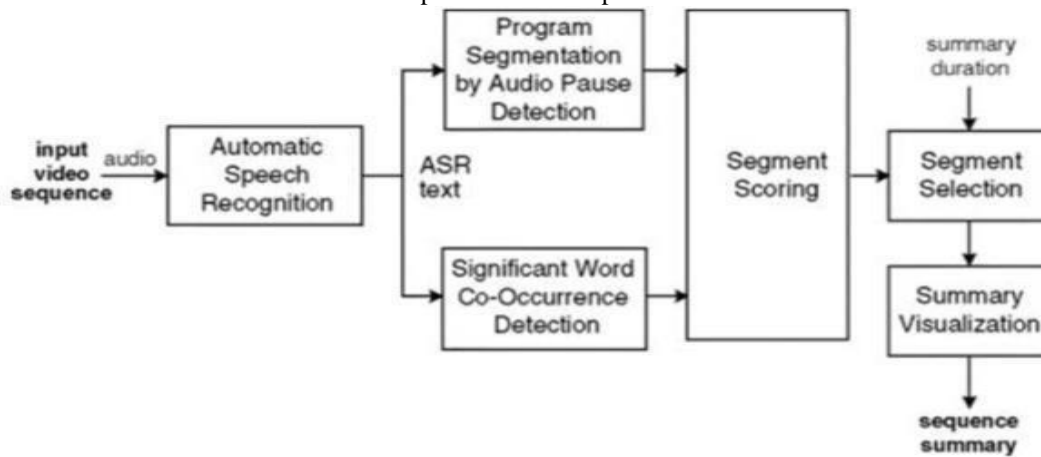

FIGURE 4.

Language translation:
A language translator is a very helpful tool that allows us to communicate in many languages by translating our language into the target language. Before there were language translation apps, it was rather challenging for people to communicate with people throughout the globe. Here, we can create our own language translation project using Python. We want to be able to translate words, phrases, and even complete paragraphs across other languages by creating a language translator. We will try to incorporate as many different languages as we can. The T-kinter Module will be used in our project to build the graphical user interface (GUI), and the google-translibrary will display arrange of languages that area part of it.

4. RESULTS AND DISCUSSION:

We will first verify that the URL is authentic. If so, we'll check to see if the transcript can be accessed. Next, we'll use the YouTube Transcript API to get the transcript and retrieve the required YouTube video. Transcript-free videos are not beings how in the feed. We process the text in order to get ready for the next action after getting the video transcript over an API. To do this, take off all punctuation, including commas and full stops, which are necessary to show the boundaries of phrases. The "punctuator" Python library can be usedto   perform   this task. The next step is to apply the text preprocessing algorithm for the transcript that was extracted. This paper can be used in the following ways: Crash Course: Students can quickly and accurately understand the material presented in this YouTube video.
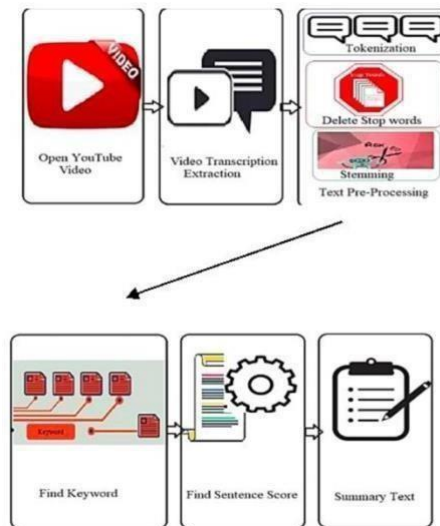

FIGURE 5.

With the introduction of online learning, students can now create summaries and their own notes using video transcripts from classes. Talks from team conferences can also be summed up using transcripts from audio and video conferences. In the context of patent research, it is conceivable to make it easier to extract significant claims from research papers and patents.
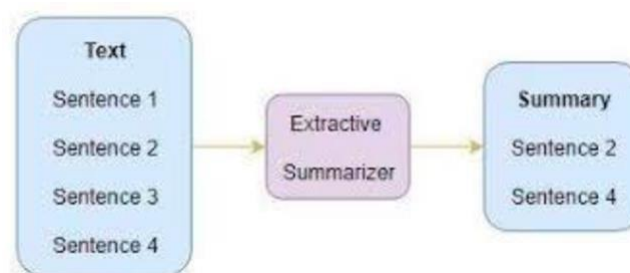

FIGURE 6.

Future scope:
Real-time summarization: For live events or news broadcasts, providing a real-time summary of happenings may be useful. Future research endeavors might focus on developing real-time summaries for the transcript.
Incorporating user feedback: By incorporating user comments, the transcript summarizer can be enhanced to produce a more comprehensive summary. Future research should focus on developing summarizers that can incorporate user input to improve the accuracy and relevance of the summary.
Image and video analysis: The transcript summarizer has the potential to generate a more precise and relevant summary by utilizing image and video

analysis technology. Subsequent research endeavors may concentrate on creating summarizers capable of scrutinizing images and videos in order to generate more pedagogical summaries.

Summarization based on user preferences: Different types of summaries may be preferred by different people. Further research could focus on developing summarizers that can generate personalized summaries based on the user's selections.

Evaluation metrics: More precise evaluation criteria are needed in order to evaluate the quality of the summaries generated by the transcript summarizer. Further research could focus on developing new evaluation standards that provide other viewpoints on the quality of the summary. In conclusion, there is a lot of space for additional research and development in the field of YouTube transcript summaries. The development of new techniques and resources could lead to more accurate and relevant summaries, which would make it easier for viewers to extract information from YouTube videos.

5. CONCLUSION

In conclusion, the user can save time by utilizing our website. We will initially view the core content of a YouTube video rather than its entire buffering waste things in order to choose which one is best forus. The user can save time and effort. They won't have to worry as much about finding the ideal YouTube video if they use our website. Additionally, text-to-speech and multilingual summary features are available on our website. We are sure that by effectively saving clients time and effort, our paper will satisfy their needs. By providing users with only the most relevant and useful information on the topics they are interested in, our approach aims to save them from having to watch lengthy movies. The time you save can be put toward learning more and experiencing new things.

REFERENCES

[1]. Alrumiah, S. S., Al-Shargabi, A. A. Educational Videos Subtitles' Summarization Using Latent Dirichlet Allocation and Length Enhancement. CMC- Computers, Materials &Continua, 70, 3 (2022).
[2]. Sangwoo Cho, Franck Dernoncourt , Tim Ganter, Trung Bui, Nedim Lipka, Walter Chang, Hailin Jin, Jonathan Brandt, Hassan Foroosh, Fei Liu,
[3]. "StreamHover: Livestream Transcript Summarization and Annotation", (2022).
[4]. S. Chopra, M. Auli, and A. M. Rush, "Abstractive sentence summarization with attentive recurrent neural networks," in Proc. Conf. North Amer. Chapter Assoc. Comput. Linguistics Hum. Lang. Technol., (2016).
[5]. Ghadage, Yogita H. and Sushama Shelke. "Speech to text Sconversion for multilingual languages." 2016 International Conference on Communication and Signal Processing (ICCSP) (2016).
[6]. Pravin Khandare, Sanket Gaikwad, Aditya Kukade, Rohit Panicker, Swaraj Thamke, "Audio Data Summarization system using Natural Language Processing," International Research Journal of Engineering and Technology (IRJET),6, 9 (2019).
[7]. S.M.Mahedy Hasan, Md. Fazle Rabbi, Arifa IslamChampa,Md.AsifZaman,"AnEffective Diabetes Prediction System Using Machine Learning Techniques", 2nd International Conference on Advanced Information and Communication Technology (ICAICT), (2020).
[8]. Prof. S. A. Aher, Hajari Ashwini M, Hase Megha S, Jadhav Snehal B, Pawar Snehal S, "Generating Subtitles Automatically for Sound in Videos," International Journal of Modern Trends in Engineering and Research (IJMTER), 3, 3 (2016).
[9]. AiswaryaKR,"AutomaticMultipleLanguage SubtitleGenerationforVideos,"International Research Journal of Engineering and Technology (IRJET), 7, 5 (2020).
[10]. Savelieva, Alexandra & Au-Yeung, Bryan & Ramani, Vasanth. Abstractive Summarization of Spokenand Written Instructionswith BERT (2020).
[11]. Patil, S. et al. "Multilingual Speech and Text Recognitionand Translationusing Image." Internationaljournalofengineeringresearch and technology 5 (2016).
[12]. S.Sah,S.Kulhare,A.Gray,S.Venugopalan, E. Prud' Hommeaux and R. Ptucha, "Semantic TextSummarizationofLong Videos,"IEEE