

Video Summarizer and Language Translator

Ganesh Kappavandla^{1*}, Rohan Vajanala², Eluri Sai Karthik³, C. Sunil Kumar⁴

^{1,2,3}Student, Department of Electronics and Computer Engineering, Sreenidhi Institute of Science and Technology, Hyderabad, India

⁴Professor, Department of Electronics and Computer Engineering, Sreenidhi Institute of Science and Technology, Hyderabad, India

Abstract: Video summarization takes a video as input and generates text from it. Video summary is a technique that selects and uses only the most significant and informative elements from a bigger video file to create a brief description for potential viewers. To translate summary material into any needed language, the language translation is utilized. The English language is well-known for its widespread use. However, many of those speakers are those who speak English as a second language. People like to speak in their native tongue. It is the language that they are most familiar with and that they utilize in their daily lives. This is why we require language translation; it will assist people in better understanding the information. This effort aims to ensure that everyone comprehends the video's information as simply and precisely as possible.

Keywords: Video summarization, language translation, content-based summarization.

1. Introduction

It is very tiring to watch a complete video to obtain information. This is problematic because there will be a lot of unnecessary information in the video, by summarizing this we can get the needed content from the video.

This project is a python code that will take any video as input and produces text from that video, and this text is summarized based on its importance.

But even after summarizing this video, few will have problems with understanding this summarized text because some people cannot understand it properly if it is not in the native language, so this code is further expanded to translate this summarized text into the preferred language.

By summarizing the video data this video can be easily interpreted and by translating the language of the video anyone and everyone can understand and access the content provided by the video even if the person could not understand the English language properly.

The purpose of video summarization is to improve the speed with which you can browse a large number of video files while also ensuring that you have the correct access to and representation of the video data. Viewers can make faster decisions on the video's functionality by looking at the output summary.

It is a known fact that the English language is widely used. But a lot of those speakers are made from users who use English as a second language. This means that many people will respond greater if they are using their mother tongue or native

language. These people can understand the English language. But until you speak the language that they know the best, their mother tongue, you can't properly communicate with them in the best way possible.

2. Literature Survey

A technique for retrieving keyframes from a long film is a video summary. They define video summarizing as a content-based recommender problem that prohibits useful content from being collected from a movie, according to this study.

To determine whether a video clip is valuable to users, a measurable deep neural network is used. They recognize scenes and events in unedited videos and make new connections between the different parts of video comprehension issues.

The impact of auditory and visual features on summarizing tasks will be investigated in this study. To reduce early-stage overfitting in the model, this technique combines data augmentation and multi-task learning.

Speech operates as a communication barrier between two people while also allowing them to express their feelings, according to one study. The process of building a communication link between robots and people is known as natural language processing. The conversion of spoken to written language is aided by speech recognition. They created a Speech Recognition model that converts speech input into text in the user's preferred language.

The current Google Speech Recognition model, which is based on some natural language processing techniques, now includes multilingual characteristics. The goal of this research is to create a speech recognition model that will allow illiterate people to communicate with computers in their native language.

3. Methodology

A. Video Summarization

Video summarization is used to create a small synopsis of the information of a larger video file by choosing and using only the most important and instructive content for possible viewers. The end synopsis is perfectly made out of a few sets of video parts or clips obtained from the initial video with some elimination process.

The purpose of video summarization is to improve the speed with which you can browse a large number of video files while also ensuring that you have the correct access to and

*Corresponding author: ganeshkappavandla@gmail.com

representation of the video data. Viewers can make faster decisions on the video's functionality by looking at the output summary. The evaluation of a summary is sometimes based on functionality studies to compute the content enlighteners and standards of a summary based on apps and viewers.

It is not possible to obtain summarized text from video directly. But summarizing the text file is simple. So video is converted into a text file and this text file is summarized. Therefore, video summarization is done in two steps.

Steps required for Video Summarization

- (a) Video to Text conversion
- (b) Text Summarization

(a) Video to Text conversion

A video file cannot be converted into a text file. But an audio file can simply convert. So video is converted into an audio file and this audio file is then converted into a text file. Therefore, video to text conversion is done in two steps.

The algorithm of video to text conversion is given below.

- Step 1: Start
- Step 2: Input = Video file
- Step 3: Extract audio from the uploaded video file
- Step 4: Using acoustic models and language models decode this audio file to obtain a text file
- Step 5: Output = Text file
- Step 6: Stop

(b) Text Summarization

Text summarization is more complex than video to text conversion. To summarize text the inputted text is segmented into sentences and this sentence is then segmented into words. Each word's value is iteratively calculated based on the word vector, word sentence, and graph model. After finding the value summarized text is generated.

The algorithm of text summarization is given below.

- Step 1: Start
- Step 2: Input = Text file
- Step 3: Segment this text into sentences
 - 1: If there is a sentence then segment the sentence and return to Step 3
 - 2: If there are no more texts then go to next
- Step 4: Segment each of these sentences into words
 - 1: If there is a sentence then segment the sentence and return to Step 4
 - 2: If there are no more texts then go to next
- Step 5: Using the relationship between words and iterative calculation of word vector and graph models find the context of the text file
- Step 6: Summarize the text file and produces summarized text
- Step 7: Output = Summarized text
- Step 8: Stop

Both steps are done simultaneously to obtain the summarized text for the given video.

B. Approaches to Text Summarization

There are two approaches to follow for summarizing text in the python language.

(a) NLTK Approach

NLTK is the abbreviation of Natural Language Tool Kit. In python programming, NLTK is a cluster of programs and libraries that is used on statistical and symbolic NLP (Natural Language Processing) for the English Language.

Algorithm of Text Summarization using NLTK

- Step 1: Start
- Step 2: Convert paragraphs to sentences
- Step 3: Text Processing
- Step 4: Tokenizing sentences
- Step 5: Finding weight frequency
- Step 6: Replacing word by weight frequency
- Step 7: Sorting sentences in descending order
- Step 8: Output = Summarized text
- Step 9: Stop

In the above-mentioned flowchart weight frequency is calculated using the formula given below.

Weight Frequency =
(Frequency/Frequency of most occurring word)

(b) spaCy Approach

spaCy is an open-source python library, which is written in Cython for NLP. It is used for building systems that extract information or understand natural language. The performance of spaCy is better compared to NLTK.

While NLTK provides access to a variety of methods, spaCy provides the most efficient method. It has the fastest and most accurate syntactic analysis of any NLP library available. It also gives you access to larger, more customizable word vectors.

SpaCy also supports deep learning processes, allowing users to integrate statistical models built using popular machine learning libraries such as TensorFlow, PyTorch, or MXNet with spaCy's machine learning library Thinc.

Algorithm of Text Summarization using spaCy

- Step 1: Start
- Step 2: Convert paragraphs to sentences
- Step 3: Text Processing
- Step 4: Tokenizing sentences
- Step 5: Filtering tokens
- Step 6: Normalizing the tokenized sentences
- Step 7: Weighing of Sentences
- Step 8: Output = Summarized text
- Step 9: Stop

C. Language Translation

It is a known fact that the English language is widely used. But a lot of those speakers are made from users who use English as a second language. This means that many people will respond greater if they are using their mother tongue or native language. These people can understand the English language. But until you speak the language that they know the best, their mother tongue, you can't properly communicate with them in the best way possible. Many people favor their mother tongue. It is the language they are most accustomed to and it uses in their day-to-day life confidently when they are using it. This is the reason why we need language translation; it will help people to communicate more successfully.

The algorithm of language translation is given below.

Step 1: Start

Step 2: Input = Summarized text

Step 3: Select the language of your choice

Step 4: Match the text to the database

1: If text matches, display translated text

2: If a text does not match, display letter by letter

Step 5: Input another text

1: If there is another text then return to Step 2

2: If there are no more texts then go to the next step

Step 6: Stop

This is language translation is done to translate any given text to a preferred language.

4. System Architecture

In this project, programming is divided into three parts

- i) Obtaining text from inputted video
- ii) Summarizing obtained text
- iii) Translating summarized text

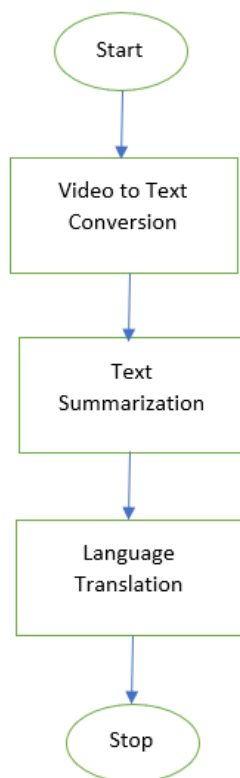


Fig. 1. Flowchart

By adding the video, the program will obtain text output from the video input. This text is summarized and then translated into the required language.

5. Results

One of the major problems during a pandemic is that almost

all learning platforms have gone online. But many students could not attend classes every day it may be because of health, family, and network issues. But it will be hard for them to prepare for examinations. So, this software is used to take videos of the lecture as input and convert them its text and then this text can be summarized for better understanding.

This can be used in many fields. Language translation can help many students to understand the content of the video more accurately. This project helps everyone to understand the content of the video as clearly and as accurately as possible.

By summarizing the video data this video can be easily interpreted and by translating the language of the video anyone and everyone can understand and access the content provided by the video even if the person could not understand the English language properly.

This project helps everyone to understand the content of the video as clearly and as accurately as possible. By summarizing the video data this video can be easily interpreted and by translating the language of the video anyone and everyone can understand and access the content provided by the video even if the person could not understand the English language properly.

6. Conclusion

The initiative of this project is to ensure that everyone comprehends the film's content completely. By summarizing the video data and translating the film's language, everyone can comprehend and access the content presented by the movie, even if they do not have a strong command of the English language. This can be used in a variety of circumstances. Many students will benefit from language translation in order to better comprehend the film's content.

References

- [1] Y. Jiang, K. Cui, B. Peng and C. Xu, Comprehensive Video Understanding: Video Summarization with Content-Based Video Recommender Design.
- [2] C. Huang and H. Wang, A Novel Key-Frames Selection Framework for Comprehensive Video Summarization.
- [3] H. Sun, R. Wang, K. Chen, M. Utiyama, E. Sumita and T. Zhao, Unsupervised Neural Machine Translation with Cross-Lingual Language Representation Agreement.
- [4] P. Choudhary, S. P. Munukutla, K. S. Rajesh and A. S. Shukla, Real time video summarization on mobile platform.
- [5] B. Kahler, B. Bacher and K. C. Jones, Language translation of web-based content.
- [6] S. Revathy and S. Nath, Android Live Text Recognition and Translation Application using Tesseract.
- [7] S. Bano, P. Jithendra, G. L. Niharika and Y. Sikhi, Speech to Text Translation enabling Multilingualism.
- [8] J. Basak, V. Luthra and S. Chaudhury, Video summarization with supervised learning.
- [9] H. Raksha, G. Namitha and N. Sejal, Action based Video Summarization.
- [10] N. Radha, Video retrieval using speech and text in video.