International Journal for Innovative Engineering and Management Research
A Peer Revieved Open Access International Journal
www.ijiemr.org

IJIEMR Transactions, online available on 5th Jan 2023. Link :http://www.ijiemr.org/downloads.php?vol=Volume-12&issue=Issue 01

DOI: 10.48047/IJIEMR/V12/ISSUE 01/07

Title **TEXT EXTRACTION FROM VIDEO CLIP**

Volume 12, ISSUE 01, Pages: 68-72

Paper Authors
CH. Swetha, D.Anusha,V. Gowthami, J. Snigdha

TEXT EXTRACTION FROM VIDEO CLIP

CH. Swetha*, D.Anusha[1],V. Gowthami[2], J. Snigdha[3]

*Asst Professor, AI Department, Vidya Jyothi Institute of Technology
[1]Asst Professor, AI Department, Vidya Jyothi Institute of Technology
[2]Student - Department of Artificial Intelligence Engineering,
[3]Student - Department of Artificial Intelligence Engineering,

Abstract

The entire definition of interaction is being redefined in today's quickly digital environment. The ever-evolving, ever-expanding realm of education is one of the areas where significant progress is being made. Due to the fact that most videos are uploaded by subject-matter experts and are very effective audio video learning resources, there is an increased demand for video tutorials for study purposes, online courses, and tutorials for entrance exams. However, occasionally the videos are too long, time-consuming, and succinctly explained for the average user. Advanced NLP is used to extract the video tutorial's audio, which is subsequently transformed to text representation. Now, we can pull text out of videos.

Key words: Speech Recognition, Moviepy, Text Extraction, MPEG (Moving Picture Experts Group) video frames.

1. Introduction

With rapid technological innovation and increased internet speed, the focus of the people are migrating away from television and toward YouTube. The primary benefit of YouTube versus television is that YouTube gives shows at the user's convenience regardless of time. There are programmes on television that are displayed at a specific fixed time, creating a temporal constraint for users. As the emphasis shifts in relation to YouTube, the article has presented a method that allows users to easily access information contained by the language in these movies. A more efficient and faster method. The suggested method would translate the video's text into editable text that will be saved in a text file. YouTube is a popular destination for news and educational videos. These videos have text, which adds information to the video and makes it more meaningful. If the text from the movies is converted to editable form, it may be saved efficiently and accessed more easily in the future. If the text from the movies is converted to editable form, it can be stored efficiently and accessed more easily in the future. After seeing the educational video, the user may not want to watch it again because he has already seen it, and reading the important points may be sufficient for him to review the topic from that video. In this scenario, the proposed method assists users in gaining access to information by transforming text in video to editable form. The editable file format is text. The biggest advantage of text file format is that it uses relatively little space when compared to video file format. Furthermore, text material can be modified if it changes in the future or if the user wishes to add any extra information to it, which is not feasible with video.

The working of the proposed system is very simple. User downloads the video form the YouTube or any other website from which he wants to extract text. This video is provided as input to the proposed system. Proposed system converts video into series of frames and applies text detection and extraction on each frame. The detected text from each frame is stored in text file.

Literature Review:

The technique that reduces character mistake rates and also eliminates noise from the characters, which significantly impairs optical character recognition, has been proposed by Datong Chen and Jean-Marc Odobez [1]. Combination edge-based text detection has been proposed by Mati Pietikainem and Oleg Okun [2] and can be used with photos that have complicated backgrounds while minimising text extraction quality degradation.The combined edge-based technique [2] was proposed by C. P. Sumati and N. Priya [3].This technique is sensitive to text orientation and skew.The system that uses entropy-based metrics has been proposed by Z. Cennekove, C. Nikou, and I. Pitas [4]. It entails comparing the colour histogram of each frame to the histogram of the following frame. When the colour histogram values of two separate photos are identical, this approach is invalid. Using border and perimeter, Priti Rege and Chanchal Chandrakar [5] have described how to separate text images from document images.Sobel operator and thresholding are used to achieve text detection. The retrieved text may be loud because text enhancement is not being performed. Text extraction using a linked component based method has been explained by Arvind and Mohamed Rafi [6]. This method requires that the text and background contrast more sharply than usual. Text identification in MPEG (Moving Picture Experts Group) video frames was described by Lifang Gu [7]. Redundancies in spatial and temporal data are decreased. Only MPEG videos can be converted using this method. Using automatic video text extraction, Baseem Bouaziz, Tarek Zlitni, and Walid Mahdi [8] explained it. It does video indexing based on content. This approach can only find static text that has been overlaid. To remove noisy blobs from the image, Punit Kumar and P. S. Puttaswamy [9] presented an approach that uses area-based filtering. When the background has more abrupt changes in intensity, this strategy fails.

Proposed System:

This system's goal is to convert a video into text so that the user may understand what is being shown without having to watch the entire thing. It mostly accepts English-language videos as input and outputs a full text description of the content of the inputted video. The language of the text is English. The proposed method operates in a relatively straightforward manner. The suggested approach divides a movie into a number of frames and uses each frame to extract and recognise text. Each frame's identified text is saved as a text file.

I. Speech Recognition:

The process of accurately translating voice data into text is known as speech recognition, commonly referred to as speech-to-text. Any programme that responds to voice commands or questions must use speech recognition. Speech identification is particularly difficult because of how quickly individuals speak, how their words are slurred together, how their tone and emphasis vary, how they speak in various dialects, and how frequently they use poor grammar. The primary advantage of speech recognition is searchability. Speech recognition is an interdisciplinary subject of computer science and computational linguistics that develops approaches and technology to enable the recognition and translation of spoken language into text by computers. It is often referred to as speech to text, computer voice recognition, or automatic speech recognition (ASR) (STT). It draws on expertise and research from the domains of computer science, linguistics, and computer engineering. Speech synthesis is the opposite process. A speaker must "train" (also known as "enrol") some speech recognition systems by reading text or a small vocabulary to the device. The accuracy of the speech recognition is improved by the system's analysis of the individual's voice and utilisation of that information. Systems without training are referred to as "speaker-independent" systems[1]. Training-based systems are referred to as "speaker dependent". Voice user interfaces such as voice dialling ("call home"), call routing ("I would like to make a collect call"), domestic appliance control, search keywords ("find a podcast where particular words were spoken"), simple data entry ("enter a credit card number"), preparation of structured documents ("create a radiology report"), identifying speaker characteristics, and so on are examples of speech recognition applications. [2]

II. Moviepy:

Using basic operations (such cuts, concatenations, and title insertions), video compositing (also known as non-linear editing), video processing, or complex effects, MoviePy is a Python module for video editing. The majority of popular video formats, including GIF, can be read and written.

III. FFMPEG:

Fast Forward MPEG is referred to as FFMPEG. It is a piece of open-source software for handling audio and video. It is command-line based and allows you to manage multimedia files, streams, and audio and video files. It's used for transcoding, basic editing, scaling, post-production effects, and standards for videos. Compliance also enables the processing of live audio and video from a source.

IV. Module description:

The implementation consists of three modules. They are the Text Module, the Audio Extraction Module, and the Text Extraction from Audio Module. The audio from the video is taken out in the first module. In the second module, the text is extracted based on the audio's language before being translated into English. The text module is the final module; in this module, the text is summarised by locating the essential terms. Only the text is extracted from the audio clips and saved as a text file.

Audio extraction module: In this module we will extract the audio from the video file or the recording by using the library called Moviepy. Now we extract the audio importing this library Moviepy.

Extraction of text from audio module: This module is the extension of the first module. Now by using Speech.

Recognition module we will convert the extracted audio clips into the text.

Text Module: In this module, the text extracted from the audio clips and all the text from the audio clips appended and finally get saved as a text file.

CONCLUSION:

In this paper we discussed our proposed method of extracting text from the video. The system is implemented in Python language. The technique can be used primarily for news and educational videos that include text-based content.

REFERENCES:

[1] Datong Chen, Jean-Marc Odobez. “Text detection and recognition in images and video frames” The Journal of The Pattern Recognition Society, 2004 pages 595-608.

[2] Matti Pietikainen and Oleg Okun. "Text extraction from grey scale page images by simple edge detectors" Machine Vision and Intelligent Systems Group.

[3] C.P.Sumathi, N.Priya "A Combined Edge-Based Text Region Extraction from Document Images" International Journal of Advanced Research in Computer Science and Software Engineering Volume 3, Issue 8, August 2013 ISSN: 2277 128X.

[4] Cerenkov, Z Greece Nikou, C. Pitas, I."Shot detection in video sequences using entropy based metrics" Proceedings. International Conference on Volume: 3 2002.

[5] Priti P. Rege Chanchal A. Chandrakar "Text-Image Separation in Document Images Using Boundary Perimeter Detection" ACEEE Int. 1. On Signal & Image Processing, Vol. 03, No. 01, Jan 2012.

[6] Arvind, Mohamed Rafi "Text Extraction from Images Using