I.J. Wireless and Microwave Technologies, **2023, 1, 27-35**
Published Online on February 8, 2023 by MECS Press (http://www.mecs-press.org/)
DOI: 10.5815/ijwmt.2023.01.03

Implementing Video OCR along with SWT Technique for Video indexing and Analysis

Paruchuru Grishman*
IARE/IT/Hyderabad, Telangana, 500043
E-mail: 18951A1228@iare.ac.in
ORCID iD: https://orcid.org/0000-0002-1493-2824
***Corresponding Author

Akula Rajitha
IARE/CSIT/Hyderabad, Telangana, 500043
E-mail: a.rajitha@iare.ac.in
ORCID iD: https://orcid.org/0000-0002-0188-1618

Mohammed Khaja Moinuddin
IARE/IT/Hyderabad, Telangana, 500043
E-mail: 18951A1249@iare.ac.in
ORCID iD: https://orcid.org/0000-0002-1727-6623

Mannava Subhramanaya Sreekar
IARE/IT/Hyderabad, Telangana, 500043
E-mail: 17951A1291@iare.ac.in
ORCID iD: https://orcid.org/0000-0002-0754-9720

Siddam Jayanth
IARE/IT/Hyderabad, Telangana, 500043
E-mail: 18951A1235@iare.ac.in
ORCID iD: https://orcid.org/0000-0001-6234-5758

Received: 04 June, 2022; Revised: 29 July, 2022; Accepted: 25 August, 2022; Published: 08 February, 2023

Abstract: The main purpose of this paper is to expand the usage of OCR (Optical character recognition) as this is only implemented over images and to extend this Video OCR is introduced in a way to help to retrieve the information from the video without playing the video. Video OCR is executed with the assistance of OpenCv2 module and PyTesseract [7] at the side of SWT approach which all pretty collectively make an ideal aggregate to offer an appropriate content from the video (i.e., Lecture video or any kind of video which has slides or text on the background of the video) [2,4].This technique is performed in a well-designed along with easy steps to provide us an correct end result of the facts from the video into textual files. In addition to this we also added Speech Recognition module within the project to support the video along with the text file. This speech delivered by the faculty (i.e., instructor/educator/teacher), or an educator will be also resulted in a text file.

Index Terms: Optical Character Recognition, Tesseract, Binarization, Python, Segmentation, Stroke Width Transform, Open CV, Video Indexing, Image Processing.

1. Introduction

OCR is merely used for image recognition to extract text from them, so the introduction to Video OCR is the main objective of this using PyTesseract and OpenCv2 to perform. The given exiting methods on OCR are about to get the text or characters from the images via using different technologies. But there is no technique or methods to get the characters from Video. So, in this paper we introduced the Video OCR with the help of different operations.

OCR stands for "Optical Character Recognition". It's a fabrication that acknowledges textual content inside a virtual picture. It's generally used to restrain textual content in oversaw cues and images. OCR software program may be used to transpose a physical paper range or a illustration into an available digital interpretation with textual content. For illustration, in case you forget about a paper train or snap with a printer, the printer will maximum possibly produce an education with a virtual picture in it. The education will be a JPG/ TIFF or PDF, still. The brand-new digital education may also nevertheless be most effective as a picture of the unique train. You also can load this scanned digital train is created, which includes the picture, into an OCR operation.

The OCR operation is on the way to arresting the textual content and converting the train to an editable textual content education software program approaches a virtual picture by chancing and spotting characters, like letters, numbers, and symbols. Some OCR software programs will really export the textual content, whilst different operations can convert the characters to editable textual content at formerly with inside the picture. Advanced OCR software program can export the length and formatting of the textual content in addition to the format of the textual content manufacturing installation on a runner.

The demerits of OCR as OCR text works efficiently with the published text only and not with handwritten text. Handwriting must be learnt by the Machine. OCR systems are expensive. Here is the need of lot of space demanded by the image produced. The quality of the image can be lost during this process. Quality of the ultimate image depends on quality of the first image. Not 100 percent accurate, there are likely to be some misapprehensions made during the system

Tesseract is an OCR machine; it's designed to use to get the textbook from the image. It has a power to assay over 100 plus languages, and it has origins in OCR ' number Python- base LSTM perpetration. Now, the rearmost fete an image which indeed has a single character. In this model there are majorly three processing way, videotape Indexing process, recognition process and the conversion of speech to textbook process. In- order to get the affair of textbook from the videotape. So, the original step is to perform videotape indexing via taking time as a parameter and the frame rate will be set to a certain point for say set the time at0.5 secs and also we get the images at every0.5 secs from the videotape these images are been elevated and encouraged to perform OCR as in the use of the PyTesseract machine returns the information on the image and this data we're subjoining into a textbook document and there goes the final result of our document which we return as an textbook from the videotape. The coming step from this module is to apply speech recognition module and returning the speech in the videotape into a textbook document.

The basic steps that comprise the model are:

1. Input Video: The input taken in the model is via accessing to file system and then the selected file should be of video file with extension of "MP4".
2. Video Slicing: Video is fragmented at every "0.5" secs and then the snap of the video has been taken.
3. Segmentation: In segmentation, the position of the character in the extracted image or snap is found out and the size of the images is cropped to each character at the template size.
4. Recognition: This works as the cropped part from the segmentation undergoes through recognition and get the data from each image.
5. STT: This step is an additional step that was different to previous, and this will be done simultaneously with the OCR as this uses speech recognition module and returns the speech of the video into a text document.

In this paper, PyTesseract, OCR and Speech Recognition techniques were implemented using various modules which are Moviepy (), Speech Recognition, Open CV2.In the section 2 there is contribution of paper, in section it is about the related work of the model and at section 4 themethodology is inaugurated.

2. Contribution of Paper

The usage of OCR is in great advantage to today's world as many of the global partners have been computerizing each document mainly the financial institutions and the repository of the government have been waiting to computerize the text from the hardcopy of document to softcopy in a easy method and a quick process. As, the demand for OCR had been rising over past three decades there is still more to develop and research on the usage of the OCR and its promising properties. Video OCR is also a recent progress from the research that this helps the recorded videos to get digitalized ina small storage and save the spaces and decrease the volume of maintenance and cost.

Following are the drawbacks of OCR:

1. OCR works only with printed text
2. Different kinds of styles of handwriting should be trained.
3. To work with OCR systems, they are expensive and also, they require huge space to process.
4. The image quality will be degraded as in the process when compared to original image

3. Related Work

The study about OCR is more of a conclusion of the results acquired by researchers from different kinds of methodologies to do a there are multiple ways to do starting from the image using binarization, noise reduction, skew correction, Thinning and skeletonization, thresholding and morphological operations. All these processes are OCR pre-processing operations which helps us by doing major of the work and reducing the challenges from OCR like scene complexity, and uneven brightness of the scanned image, rotations of the characters, blurriness and pixel degradation and a major drawback of not useful for multilingual usages.

As, the studies says that after performing any of the pre-processing state then the segmentation phase begins that helps us to isolate a character image like a cropped image of each character from the image and then the implementation of OCR does. Later the Normalization phase in this if there are any unwanted gaps or missing information is filled and elimination of unwanted data is performed.

The existing method of the OCR main purpose of to extract the information from the images.

As these days the major work of the documents is digitalized. For, example if you want to write a grocery list you will write it in a note from your phone and this OCR also an advantage to digitalize the text from the images or the handwritten text on the slips or reminders.

4. Methodology

In the existing system we have added two of new implementations:

4.1. Stroke width Transform

The SWT is a local operator which calculates for each pixel the width of most likely stroke containing the pixel. This will be done by first grouping pixels with similar stroke width, and then applying several rules to distinguish the letter candidates. Since single letters won't appear in images, we will group closely positioned letter candidates into regions of text. Optical Character Recognition (OCR) is used to convert the printed text into machine encoded text and returns it as output.

4.2. Using of Speech Recognition

In this we add the module of Speech Recognition to convert the video speech to text and make a use of it with resulting

4.3. Working

Users give an input video via browsing the system i.e. file explorer and then select the video file with "mp4" as an extension and this file will undergo the process of OCR with help of tesseract engine and then this provides with the result of text from the video by processing this video into image and then later the text will be extracted from the image then it will deployed in an text document after the video completed executing we will get the final text document of the output from the video to image to text. Next, we have one more result from the video the speech to text performance here we get the text from the speech that is lectured or explained in the video this will be converted into a text document. So, finally we will be getting a resultant of two text documents.

OCR optical character recognition is an essential technology that has been in a part of life over three to four decades, today the most advanced OCR software produces transcript of most forms of handwritten and these can be turned into computer printed formatted text which is very helpful while converting from documents to online[5]. While using this OCR we have used it a PyTesseract which is tesseract engine. Python test practice and optical character recognition tool for Python this recognizes the text embedded in images. There is another module that is used is PIL pillow this is mainly an image module which provides a class with the same name which is used represent as a PIL image this helps us to 31 load an image and rotate an image or display an image and create thumbnails for an image and there are many functions and many methods that has been used using pillow this there are some lazy operations that can be done using this function identifies the file but the file remains open and actual data image data is not read from the file until that we try to process the actual data and there are many file manipulation methods like file.seek(), file.tell() ,file.read().

4.3.1 System Architecture

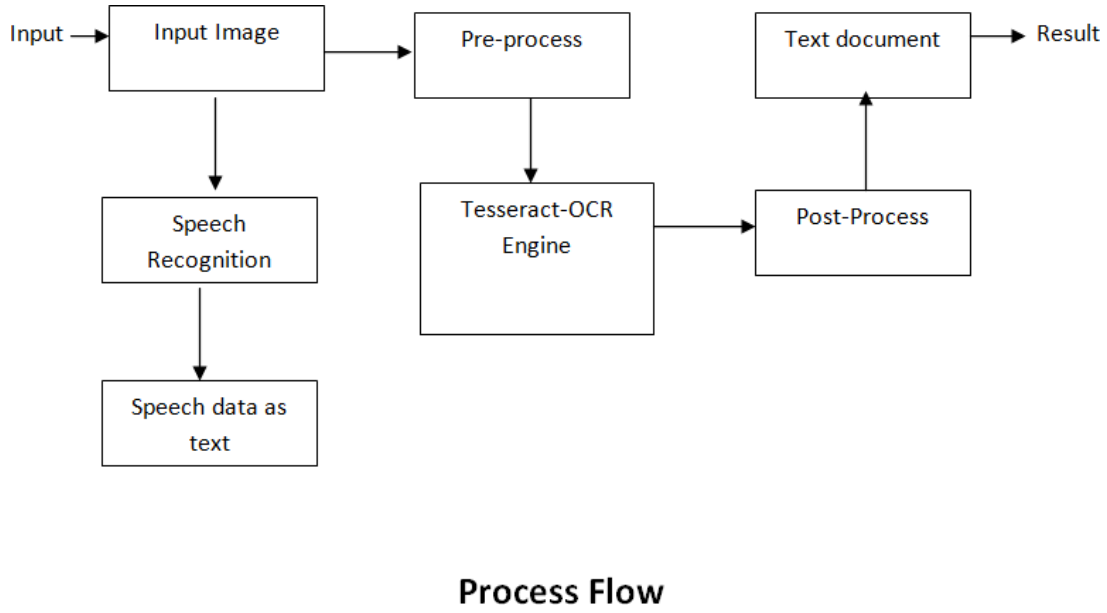

Fig. 1. System Architecture

For implementing the project, we have defined four parts of the whole theme and they were Front, Video, OCR, STT these have been a great part to develop and implement the software. The front includes the part of python GUI which imports the package of tkinter, and this helps us to choose the video file and give the input to the software. The Video this is where the frames/snaps will be taken from the video according to the prescribed seconds or frames per second .Next, the OCR here is the place the magic happens in this part the images undergo through tesseract engine and complete the process of extracting the text from the image and stores in a text file ,then STT i.e. Speech to Text, here we convert the video to audio and then to text and this text will be displayed in a text document.

The overview of working model is it starts with two functionalities such that at first work the Speech to text of the Video is done using Moviepy module and Speech recognition module, now the second work is about Video OCR [1,12,6] in this workflow it contains of six phases coming through at initial the video indexing is done and an array of images is developed from the video and then sent to pre- processing here the elimination of noise images and it enhances the image quality and make it ready to next phases now the second phases are segmentation phase in this the image is divided by each line of text into image and then the line is divided to each word into image and then each letter into image from word then the letters were cropped and then the text extracted and next at third step the normalization is performed in this the data is been cleansed and removing of unwanted information is done. Fourth phase is about classification with this the pattern of text is designed and segregated with relevant content and then the last phase is about the postprocessing this tries to clean the record in a specific sense.

Here the implementation of OCR has been done through tesseract as the extraction of content present in image is being withdrawn and added to a text file, the process goes as like when the video indexing has been done and then the images have been extracted and these images will be sent to OCR and the extraction has been done using OCR using PyTesseract engine to collect the data from an image and here is a text file document which has no content in it as an empty text document where the function appends the data or the content from the image to text on the text document and after completion of execution of all the images which have been sent by the video the file would be closed end the function exits.

Later, after completion of the execution of OCR over the video and then the process of the STT i.e., Speech to Text is going to initialize and this would be done by processing the video to audio as the file will changed to ".wav" extension from ".mp4" with a user defined name of the audio file.

Now, this audio file will undergo the speech recognition module and convert the speech to text and this text will be appended in a text document.

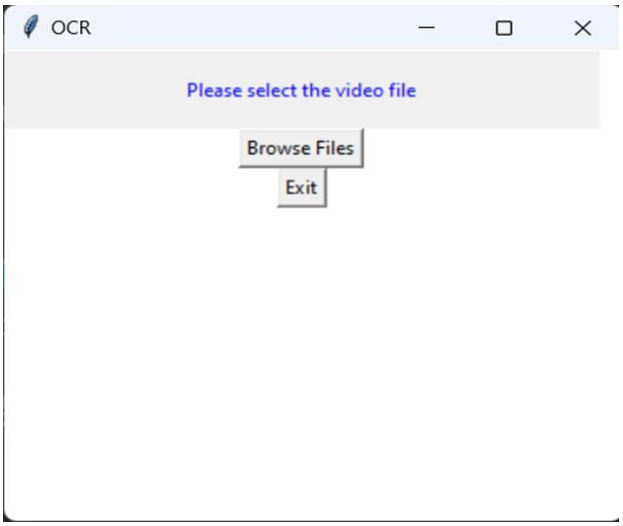

Fig. 2. Front-end screen

Moviepy() is a Python module that is used for video editing most probably it is used for video cropping like cuts concatenations and insertions of text or titles and composing of video like nonlinear editing and video processing and also to create effects and it can be used for reading of the video [12] and writing are like appending the video and the most common video formats and also to convert the videos into GIFs it is very easy to use moviepy() module as it is can use it to fetch all the image files and also we can read the image data and we can store the images into a list or an array and we can even create a video writer object and we can save the images and whatever formatted name text we needed and we can even add and delete audios from this video and there are some animation effects that can be used and we can be used this as to implement text or video and also to define frames in the video it has automatic video editing options which once the program executes it can be used in many ways.

Here the implementation of OCR has been done through tesseract as the extraction of content present in image is being withdrawn and added to a text file, the process goes as like when the video indexing has been done and then the images have been extracted and these images will be sent to OCR and the extraction has been done using OCR using PyTesseract engine to collect the data from an image and here is a text file document which has no content in it as an empty text document where the function appends the data or the content from the image to text on the text document and after completion of execution of all the images which have been sent by the video the file would be closed end the function exits.

The executing of the project model will be held in a queue method after completion of the STT process then the process of OCR will be initiated and then we receive a status as "Video Executing completed" this indicates that the process of STT and OCR has been successfully completed and we can perform another operation or safe to close the window.

```
MoviePy - Writing audio in converted.wav
chunk:   0%|          | 0/2695 [00:00<?, ?it/s, now=None]chunk:  11%|          |
 283/2695 [00:00<00:00, 2809.23it/s, now=None]chunk:  25%|          | 663/2695 [
00:00<00:00, 3349.25it/s, now=None]chunk:  37%|          | 1001/2695 [00:00<00:0
0, 3232.29it/s, now=None]chunk:  49%|          | 1326/2695 [00:00<00:00, 3026.42
it/s, now=None]chunk:  61%|          | 1632/2695 [00:00<00:00, 2706.99it/s, now=
None]chunk:  71%|          | 1915/2695 [00:00<00:00, 2484.36it/s, now=None]chunk
:  81%|          | 2170/2695 [00:00<00:00, 2219.01it/s, now=None]chunk:  89%|
          | 2399/2695 [00:01<00:00, 1924.76it/s, now=None]chunk:  97%|          | 2
601/2695 [00:01<00:00, 1682.12it/s, now=None]
```

Fig. 3. Converting of video to audio file in console screen

The above screen indicates that the process of converting the video file to audio file is ongoing and this audio is writing into a file name "converted" with an extension of ".wav".

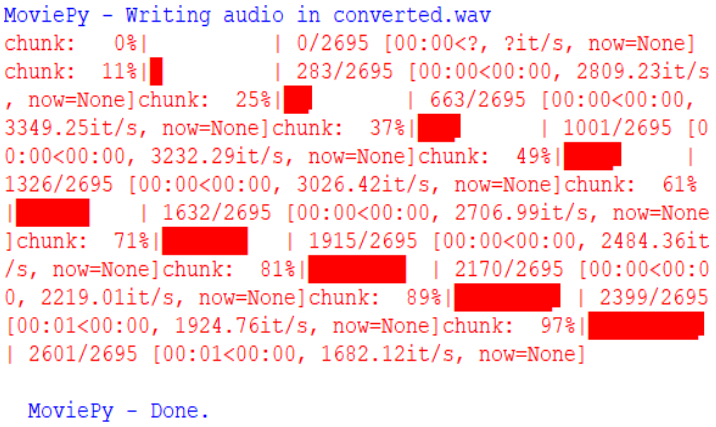

Fig. 4. Converting of video to audio file is done in console screen

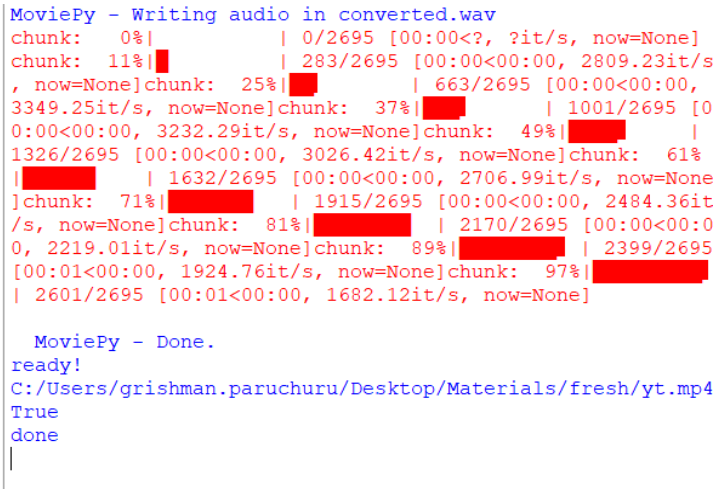

Fig. 5. OCR of the video file is done in console screen

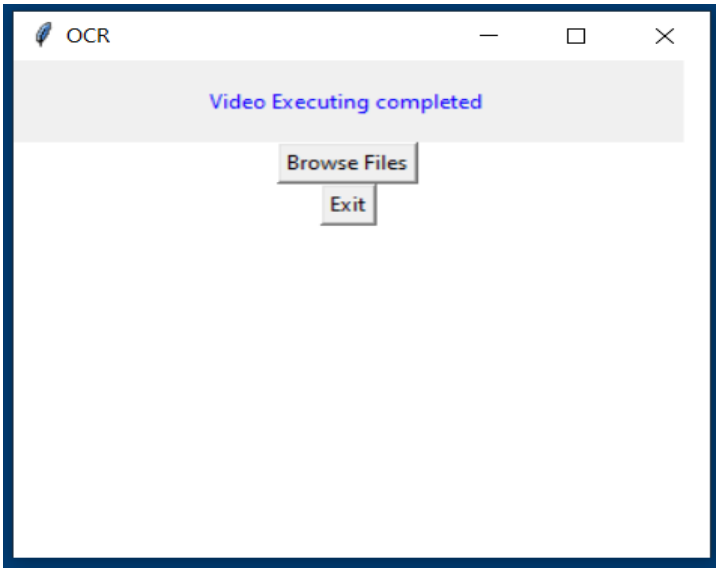

Fig. 6. Video Executing is Success on front-end screen

The fig.4 shows that the speech to text is done, and the OCR flow is done in fig.5 all on the console screen and the fig.6 employs on the front-end that the executing of the video is completed successfully.

5. Results and Discussion

After brief analysis the obtained text document of the content in video with supporting files and we managed to maintain a user-friendly environment to understand the functionality even to a lowbrow person and the process of OCR is easily defined with separate methods so that can use it anytime at our convenience and from concluding our proposed system we represent that by using of OCR from PyTesseract from OpenCV and the module of speech recognition we get the data from the video with all the requirements of the data in it as in two documents one consist of the ppt or screen information in it and the other speech given by the instructor.

As the part of Software Development life Cycle a unit testing is done at a certain level that both manual field testing and detailed functional testing are in working successfully. Indicators of success : there must be no errors in any of the fields and selection of the video file is to be when clicked on the designated button. This includes the entry screen and responses of the process flow in the console. Aspects that will be put to the test. It's important to make sure that all the video file is given correctly. There should be no room for duplicate entries. The selected video file must undergo through the process of Speech to text and the OCR mechanism with this it can be taken to high level end and make sure it can handle the future endorsements and will be a helpful to next process.

The resultant output can be declared in text documents in concluding that from the above concepts where the file input is being taken using front end and the extraction of the images from the video is done at back end using OCR and test tracked engine and the main process of conversion of image to text is done at best tesseract and the speech to text module where the speech is converted to text and place it in a text document using speech recognition and also movie Pi which helps us to convert a video file to an audio file and their fight defining this audio file into text which will be appended in a text document and at last the result would be two text to documents as a resultant which have the information of the content displayed on the slides of a presentation of the video and the other text document contains the information about the lecturer explaining the topic in the video.

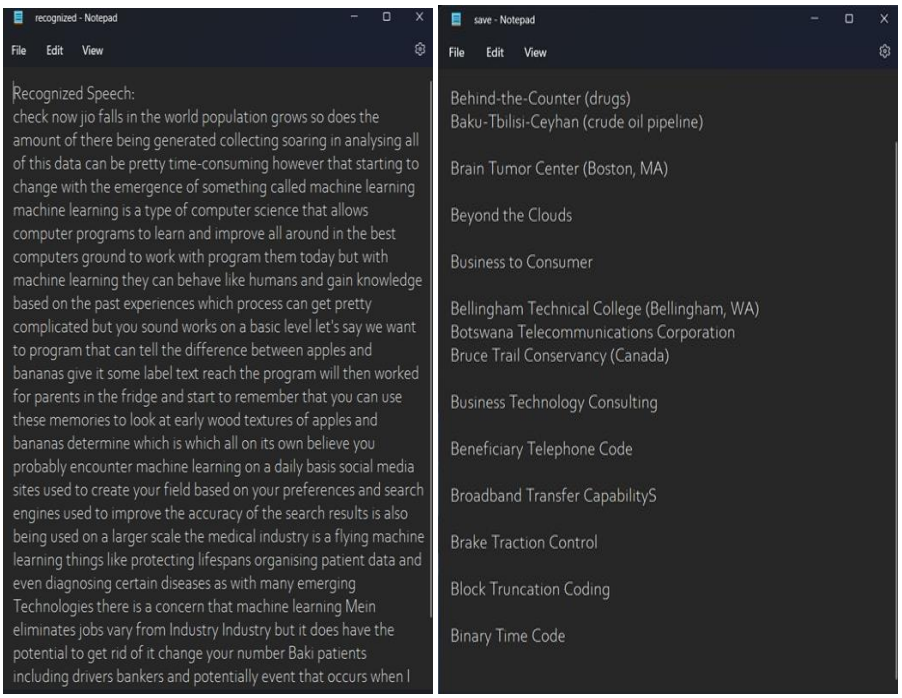

Fig. 7. Result of the two text documents

6. Conclusion and Future Scope

Optical Character Recognition has been round for the ultimate 80 years. In any case, at first, items that understand optical characters have been for the maximum component advanced via way of means of large innovation organizations. Improvement of AI and profound gaining knowledge of has empowered man or woman experts to foster calculations and methods [6], that can understand manually written authentic copies with extra noteworthy precision From Concluding our proposed system we represent that by using of OCR Engine from PyTesseract from OpenCV this and the module of speech recognition we get the data from the video with all the requirements of the data in it as in two documents one consist of the Power point presentation or screen information in it and the other speech given by the instructor. This has been helpful to many students over past 6 months as it gives us the transcript of the video file and the text from the ppt without even opening the video file or the lecture video.

After observation of this software project the decision can be hold up to a single file at each time of performing OCR at this level it takes time to do a ton of video files. So, the future build would be to take an input of a folder or selecting multiple files at a time and this would be a great advancement. As for the model this can be customized by any other user as this easy to adapt and utilize. Any part of this software is customizable and can be developed at the hands of any software engineer and the used technology OCR there is no certain limitation that it could work only to limits it has been used since 1990's so there never has been a struggle to implement this, and it has made many lives easier during the conversion of offline work of docs to online.

References

[1] Karez Abdulwahhab Hamad,Mehmet Kaya :A Detailed Analysis of Optical Character Recognition Technology.
[2] Jay Dilipbhai Thanki, Priyank Dineshbhai Davda, Dr. Priya Swaminarayan :A Review on OCR Technology
[3] Jamshed Memon, Maira Sami, Rizwan Ahmed Khan,Mueen Uddin: Handwritten Optical Character Recognition (OCR): A Comprehensive Systematic Literature Review (SLR)
[4] Boris Epshtein, Eyal Ofek,Yonatan Wexler :Detecting Text in Natural Scenes with Stroke Width Transform.
[5] Mahitha G, Surabhi K, Rahul Kumar: Enhanced Stroke Width Transform to Detect Text Regions in NaturalScene Images.
[6] Chirag Patel, Atul Patel, Dharmendra Patel :Optical Character Recognition by Open-source OCR Tool Tesseract.
[7] Muskan Chawala , Rachna Jain, Preeti Nagrath :Implementation of Tesseract Algorithm to Extract Text from Different Images.
[8] Minghui Liao, Zhaoyi Wan , Cong Yao, Kai Chen, Xiang Bai-Real-time Scene Text Detection withDifferentiable Binarization
[9] Karishma Tyagi, Vedant Rastogi Survey on Character Recognition using OCR Techniques
[10] Shiravale, Sankirti & Kamade, P. (2011). Video OCR for Video Indexing. International Journal of Engineering and Technology. 3. 10.7763/IJET.2011.V3.239.
[11] Avinash Verma, Deepak Kumar Singh,"Text Deblurring Using OCR Word Confidence", International Journal of Image, Graphics and Signal Processing(IJIGSP), Vol.9, No.1, pp.33-40, 2017.DOI: 10.5815/ijigsp.2017.01.05
[12] C. Gonzalez Richard E. Woods "Digital Image Processing" Book Third Edition Rafael Interactive Pearson International Edition prepared by Pearson Education PEARSON Prentice Hall.
[13] Shamik Tiwari, V. P. Shukla, and A. K. Singh "Review of Motion Blur Estimation Techniques" Journal of Image and Graphics Vol. 1, No. 4, December 2013.
[14] Kishore R. Bhagat,Puran Gour "Novel Approach to Estimate Motion Blur Kernel Parameters and Comparative Study of Restoration Techniques"International Journal of Computer Applications (0975 – 8887) Volume 72– No.17, June 2013.
[15] Nam-Yong Lee "Block-iterative Richardson-Lucy methods for image deblurring" Lee EURASIP Journal on Image and Video Processing (2015) Springer 2015:14 DOI 10.1186/s13

Authors' Profiles

Paruchuru Grishman born in India 2001 has Done B.Tech in information Technology in Institute of Aeronautical Engineering which is Affiliated to Jawaharlal Nehru Technological University Hyderabad. His area of interest is Image Processing, Machine Learning and Optical character Recognition. He is currently working on convolutional neural network.

Mohammed Khaja Moinuddin born in India 1999 has completed B.Tech in Information Technology from Institute of Aeronautical Engineering, Hyderabad. His area of interest includes Image processing and Optical character Recognition.

Mannava Subhramanaya Sreekar born in India 1998 is graduating from B.Tech in Information Technology from Institute of Aeronautical Engineering, Hyderabad. His area of interest includes Image processing and Optical character Recognition.

Jayanth Siddam born in India 2000 has completed B.Tech in Information Technology from Institute of Aeronautical Engineering , Hyderabad. His area of interest includes Machine learning and DSA.

Akula Rajitha born in India 1992 has done B. Tech in Computer Science and Engineering VCWE. MTech in Software Engineering from JBIET. Her Area of Interest is Image Processing and Internet of Things (IOT). Her research was on IOT and Architectural based Resilience Modelling of Multi Cloud Computing Systems

How to cite this paper: Paruchuru Grishman, Akula Rajitha, Mohammed Khaja Moinuddin, Mannava Subhramanaya Sreekar, Siddam Jayanth, "Implementing Video OCR along with SWT Technique for Video indexing and Analysis", International Journal of Wireless and Microwave Technologies(IJWMT), Vol.13, No.1, pp. 27-35, 2023. DOI:10.5815/ijwmt.2023.01.03