IVP_MINI_PROJECT

Text recognition in images and converting recognized text to speech

BY 20CS01030

BHAVYA SRI SEELAM

ABSTRACT:

In recent years we have developed our technology to an extent that we can perform the unexpected with ease. Technology makes our daily life easier. Starting from eBooks to online learning, in every industry digitalization is overpowering the flow of information. When we come across the concept of eBooks, we also need to focus upon the people who are visually impaired. For this we came across the concept of audio books and podcasts. To put a cherry on top, we present our research work which can help to convert any written text to text file and further present an audio file reading the given text aloud. This will not only help the visually impaired, but will also be helpful for the illiterate who now will be able to read the text with the use of the proposed methodology. Here we will be performing multiple processes on an image to extract the text and further convert that text to audio which will read out aloud the provided text. Using Optical Character Recognition and Text To speech conversion we are presenting our research work to implement the conversion of text in image to speech.

(I) INTRODUCTION:

In recent years technology is developed faster than ever, it is the era of digitalization where Mobile Phones and other smart gadgets are the main source of communication and interaction. We make calls, texts and voice notes to communicate easily. Verbal communication is considered the best way of conveying and expressing ourselves in an effective manner. To help visually impaired, text to speech (TTS)[1] was first introduced where AI generated voice would read the text to the user. In this paper work we will look at converting text to speech using Optical Character Recognition (OCR) and Robotic Programming Automation (RPA).

OCR is widely used to extract texts from images and process the images.[9] We can develop an effective system to help the visually impaired and the illiterate to help them read banners, text from pamphlets, books and their surroundings. This research paper will mainly have 3 steps. Firstly, we get an input through camera or pre saved image file which is processed by OCR (UI Path's library is used) [3]. Secondly, the output of OCR is processed by our robot which further processes the received text and convert it into audio. Lastly, the processed audio is presented to the user.

Here RPA is a software technology that makes it easy to build, deploy and manage software robots that replace the human involvement in redundant tasks [5]. Just like humans RPA bots can simulate humans and give their inputs and keystrokes or navigate pages, identify the data and extract it. Using RPA, we can perform wide range of pre-defined actions faster than humans and is highly accurate with least chances of human error. RPA is used in this process to manage all the working in this research. With the help of RPA, we take input from

the user and provide it to OCR module. The output of OCR is managed by RPA and it further on passes the text file to the TTS engine. The final audio is received by our bot from TTS engine and is further saved or presented to the user. RPA provides an environment in which everything is processed. We use UI Path to implement RPA in this research work.

1.1 OBJECTIVE

- To effectively convert text to speech.
- To enrich the audio output.
- To enhance the image processing so as to get better and accurate results.

(II) METHODOLOGY

The proposed model can be broken down into two major modules, image processing and audio processing module. The image processing module gets an image which is directly captured or which is previously saved in the device, the text from the received image is extracted. The voice processing part, the previously extracted text is processed and converted into high quality audio output.

Firstly, the image processing module (OCR) processes .jpeg or .jpg and form a text file (.txt) having the extracted text. Optical Character Recognition can recognize the characters through the AI optical mechanism. Before providing the image to OCR it is converted to binary image for higher image recognition accuracy.

This image conversion is done by UI Path's software as shown in fig 2.1.

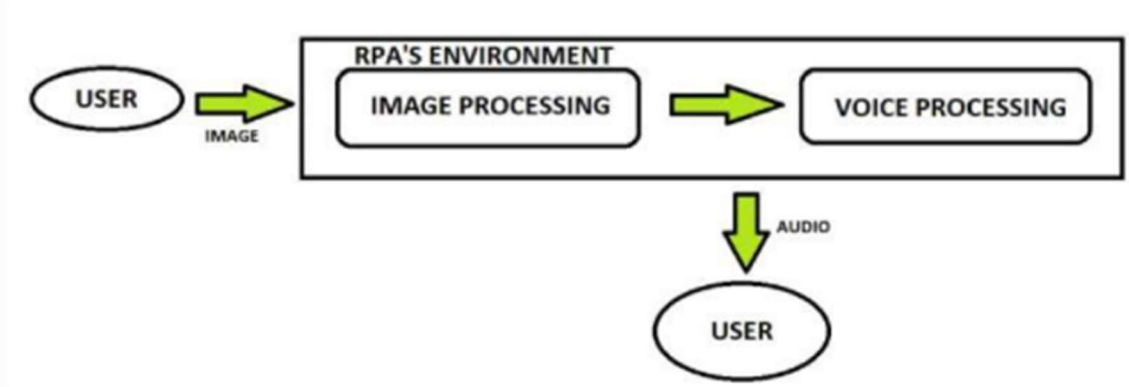


Fig. 2.1

(A) IMAGE PROCESSING

Our display is made up of pixels of distinct colours. Higher the number of pixels higher the resolution. Our devices showcase these pixels in a given structure so as to display the resultant images that we see. This phase of our research can be primarily be divided into three steps as shown in fig. 2.2. Image pre-processing, image is pre-processed before OCR

works on the image. The OCR works on the processed images and extracts the text from the given image. To avoid any error, lexical analysis is done to finalize the received text. Once all this is done, final text file is generated.

fig. 2.2

STEP-1: Pre-processing of Image

Before we extract the text from the image, it is processed and edited a few times to make the extraction easier and more successful. This is pre-processing of the image and this is done in multiple ways depending on the software we use for preprocessing.

Binarization: The received picture is converted into black and white image. In this image processing the main goal is to put all the secondary data in the background with white color and highlight all the textual data in black. This will help the OCR system to be more precise and swifter.

Deskewing: The binarized image is received and it requires further processing. All the extracted text may- may not be completely aligned. It's here when all the collected text is sorted and aligned horizontally. The mis-aligned text is rotated and aligned horizontally.

Despeckle: The received image may have noise that will make it hard for OCR to process the image. The job is to smoothen the picture and remove the noise for better recognition of the text.

Line Removal: All the lines or boxes or structures that are not text or characters are removed. This makes it easier for OCR as now it will not get confused between fake lines and characters. This helps especially in the case of document or table text extraction.

Zoning: Here the image is broken down into zones where each zone represents a paragraph or a column. This helps in distinguishing whether the text belong to a column or it’s in same line.

STEP-2: OCR Image Processing

In this step OCR does its work. At first it differentiates between character and line spacing. This works becomes a lot easier when zoning is done properly. Each line or zone of text is handled one by one by OCR

Tokenization is a process in each empty space between text is marked as a token and each token is considered as a character. OCR finds these spaces and recognizes different words.

Once all the characters are tokenized, the OCR use its techniques to identify what characters or tokens stand for. It uses the following techniques: - 1- Feature Extraction:

There are some predefined rules that guide OCR for this process. These rules describe which character is getting represented. Like a single horizontal line is likely to be 'I' or two equal horizontal lines connected diagonally may standing for capital 'N'.

2- Pattern Recognition:

Glyphs- any kind of purposeful mark, such as a simple vertical line incised on a building, a single letter in a script, or a carved symbol or any alphabet.

The earlier received tokens are now compared with the known glyphs. The glyphs include all punctuation marks, numbers, alphabets and special characters. The closest match is selected. This is often called as matrix matching as the character is divided into a matrix and each pixel of a matrix (of the token) is matched with the pixels of glyphs. [2]

The only drawback is that the font style and size should be similar to that of glyphs. Due to this, we cannot work with any handwritten text. At the same time, if the token's font is known then this process is faster and more accurate.

STEP-3: Image post-processing

Lexicon is collection of all the words and phrases used in a particular language or subject. All the received words are compared with a list of approved words called lexicon. The words that don't match are replaced with the nearest match. This can really help in distinguishing between capital 'I' and small 'L" or between zero and capital 'O'.

(B) VOICE PROCESSING:

The image processing part provides us a text file having the extracted text. We are just half way through the path. This text from text file is copied and is supplied to Text to Speech Synthesizer (TTS) which processes the text and reads it aloud. This voice note is downloaded and shared with the user. This part is completely done with RPA. A bot is created that will perform all these steps and give an audio output to the user.

TTS is a computer-based system that readout the text provided to it. Speech is based on natural pronunciation of every note and put together to form a word and further on sentences. This speech can be received in speaker or headphones. This file can even be saved for repeated playback.

(C) RESULTS AND DISCUSSIONS:

The recent development in technology have made everything easier and accessible for the users. Text can be converted into audio with the help of the described concept. The user can provide us an image through internal storage or camera directly. Our bot will do the rest. RPA provides an environment to everything for its working. UI Path is used to provide all the RPA's functionalities (as shown in fig 3.1).

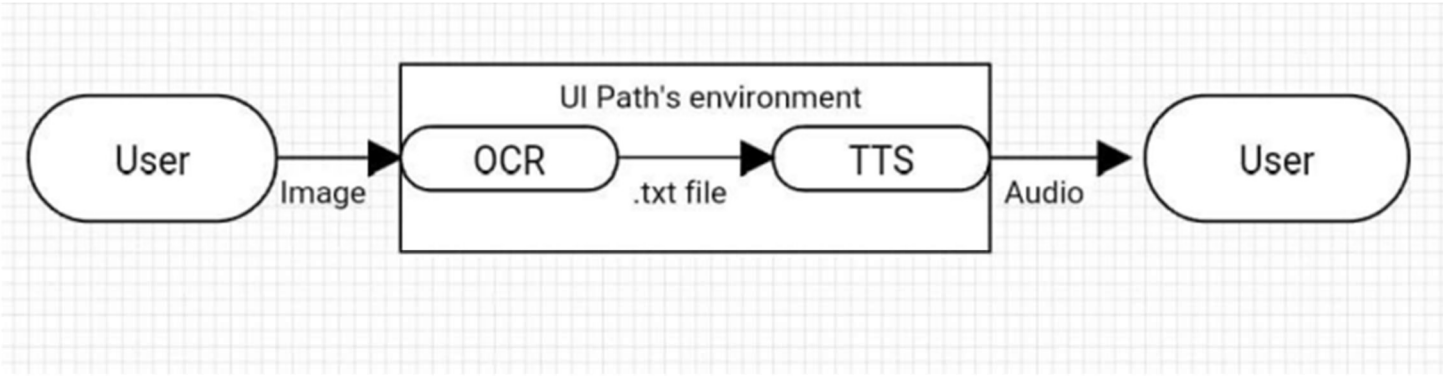


fig 3.1

The OCR receives the preprocessed image. The image is processed multiple time before actual text extraction is done. The image preprocessing is described in the fig. 3.2.

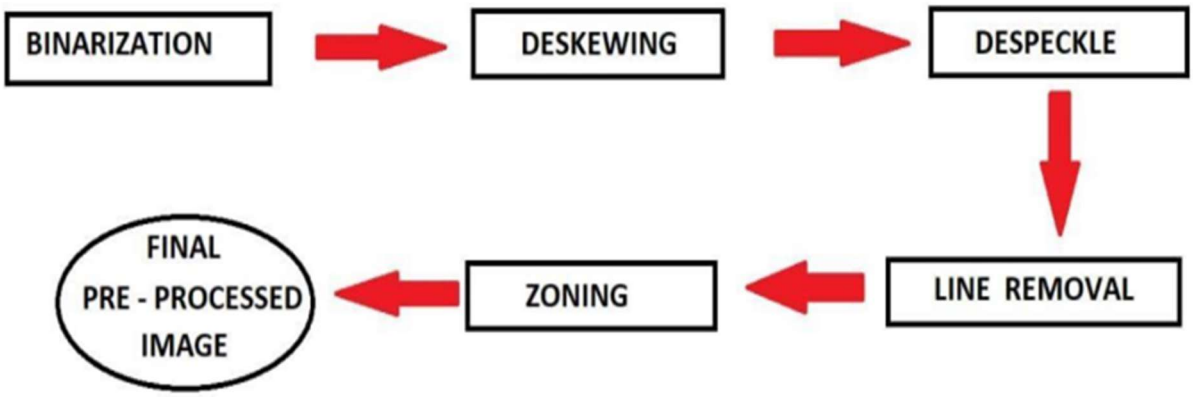
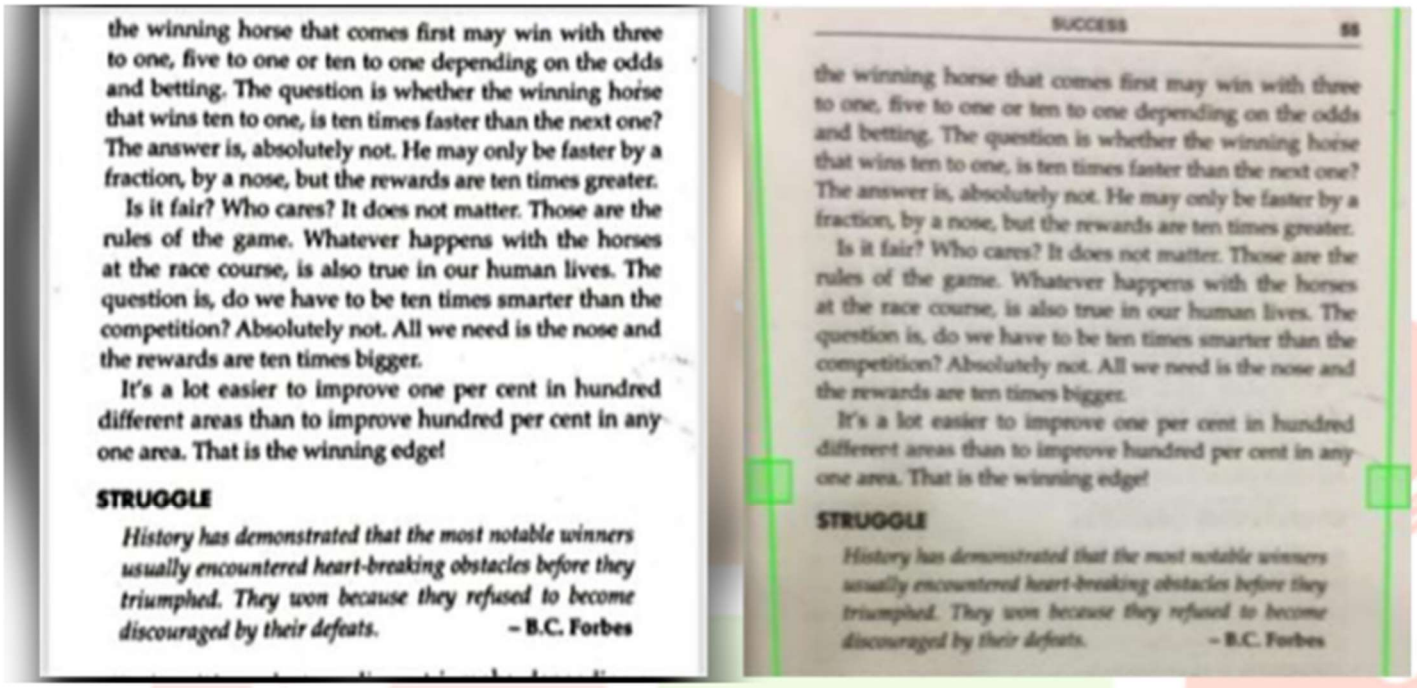


fig. 3.2.

The image preprocessing starts with Binarization of the image. The image is processed and black and white image is generated as depicted in fig 3.3.


the winning horse that comes first may win with three to one, five to one or ten to one depending on the odds and betting. The question is whether the winning horse that wins ten to one, is ten times faster than the next one? The answer is, absolutely not. He may only be faster by a fraction, by a nose, but the rewards are ten times greater.

Is it fair? Who cares? It does not matter. Those are the rules of the game. Whatever happens with the horses at the race course, is also true in our human lives. The question is, do we have to be ten times smarter than the competition? Absolutely not. All we need is the nose and the rewards are ten times bigger.

It's a lot easier to improve one per cent in hundred different areas than to improve hundred per cent in any one area. That is the winning edge!

STRUGGLE

History has demonstrated that the most notable winners usually encountered heart-breaking obstacles before they triumphed. They won because they refused to become discouraged by their defeats. – **B.C. Forbes**

SUCCESS 55

the winning horse that comes first may win with three to one, five to one or ten to one depending on the odds and betting. The question is whether the winning horse that wins ten to one, is ten times faster than the next one? The answer is, absolutely not. He may only be faster by a fraction, by a nose, but the rewards are ten times greater.

Is it fair? Who cares? It does not matter. Those are the rules of the game. Whatever happens with the horses at the race course, is also true in our human lives. The question is, do we have to be ten times smarter than the competition? Absolutely not. All we need is the nose and the rewards are ten times bigger.

It's a lot easier to improve one per cent in hundred different areas than to improve hundred per cent in any one area. That is the winning edge!

STRUGGLE

History has demonstrated that the most notable winners usually encountered heart-breaking obstacles before they triumphed. They won because they refused to become discouraged by their defeats. – **B.C. Forbes**

fig 3.3

The mis-aligned text is aligned by the process known as deskewing, shown in fig.3.4. [7]

fig.3.4

Next the noise from the image is filtered out, this process is known as despeckle as shown in fig 3.5.

fig 3.5

After these processes, the irrelevant lines are removed from the image so to make OCR's output the most accurate. This process if known as line removing.

After all of the image processing the image is divided into zones that divide the image in different columns and paragraphs as depicted in fig. 3.6. [8]

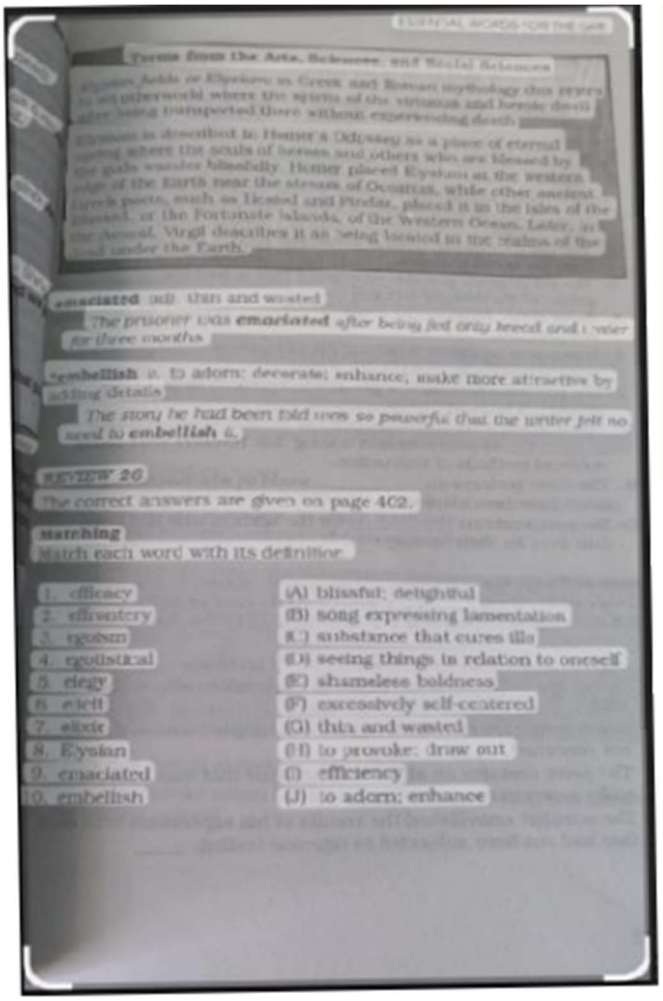


fig. 3.6

After all the work is done, the processed picture is provided to OCR module for text extraction.

Text file will be generated and our TTS engine will work and the voice output will be generated.

(D) CONCLUSION AND FUTURE SCOPE:

Through this research the complete working of text to speech conversion is conducted using RPA. We have used UI Path to implement RPA. UI Path will provide an environment for image processing through OCR and voice processing with TTS engine. This helps in reading aloud the text which is presented to the software. This is extremely helpful for visually impaired or illiterate people. For further enhancement we can translate the given text into different languages [6,10] which can further enhance the quality and functionality of the software.