e-ISSN: 2582-5208
International Research Journal of Modernization in Engineering Technology and Science
(Peer-Reviewed, Open Access, Fully Refereed International Journal)
Volume:05/Issue:05/May-2023 Impact Factor- 7.868 www.irjmets.com

IMAGE2SPEECH RECOGNIZE TEXT IN IMAGES AND CONVERT ACKNOWLEDGED TEXT TO SPEECH

Ajay Pal[*1], Vasu Chaudhary[*2], Rahul Singh[*3], Anmol Tyagi[*4]

[*1,2,3,4]Dept. Of Computer Science And Information Technology Engineering

RDEC, Ghaziabad, India.

DOI : https://www.doi.org/10.56726/IRJMETS40559

ABSTRACT

The objective of this research project is to create a tool that can transform written material into an audio format, thereby enhancing accessibility for people who are visually impaired or illiterate. The tool will employ Optical Character Recognition (OCR) to recognize and extract text from images, which will then be converted into a written document. Subsequently, the written document will undergo Text-to-Speech (TTS) conversion, resulting in an audio file. To achieve this goal, the system will execute a series of image processing operations to extract the text, which will then be transformedinto speech.

The proposed research has the potential to bring significant benefits to individuals withvisual impairments or limited literacy skills, as it will enable them to access and read textcontained in images using the developed system. This innovative methodology will helpbridge the accessibility gap for people with disabilities, providing them with greater access to information in an increasingly digital world. Additionally, the findings of this research may prove valuable for scholars and developers working in the fields of text recognition and speech synthesis.

Keywords: OCR, Text Extraction, Image Processing, Text To-Speech, Voice Processing.

I. INTRODUCTION

In recent years, technological advancements have rapidly transformed our society, leading to an age of digitalization where mobile phones and other smart devices have become integral to our daily communication and interactions. Verbal communication is often considered the most effective way to express ourselves, and advancements in text- to-speech (TTS) technology have enabled visually impaired individuals to access information with the help of AI-generated voices. In this study, we will explore the potential of using Optical Character Recognition (OCR) and Robotic Programming Automation (RPA) for text-to-speech conversion. By harnessing OCR's ability to extract and process text from images and RPA's automated programming capabilities, we can develop an effective system to assist visually impaired and illiterate individuals in reading banners, pamphlets, books, and their surroundings.

This research paper is organized into three main sections. Firstly, the system receives input from a camera or a previously saved image file, which is then analyzed using Optical Character Recognition (OCR) through UI Path's library. Secondly, the output from OCR is processed by our software robot, which converts the received text into an audio format. Finally, the processed audio is presented to the user.

Robotic Process Automation (RPA) is a software technology that enables the easy creation, deployment, and management of software robots to automate repetitive tasks that would typically require human intervention. RPA bots can simulate human inputs, navigate web pages, identify data, and extract it, thus minimizing the likelihood of human error and speeding up task execution. In this research project, RPA is utilized to manage all the processes involved. The RPA bot takes input from the user and transmits it to the OCR module. The RPA software manages the output of OCR and transfers the text file to the Text-to-Speech (TTS) engine. The final audio output from the TTS engine is then received and saved or displayed to the user by our bot. The use of RPAstreamlines the process, optimizing its efficiency. UI Path is employed to implement RPAin this study.

1.1 OBJECTIVE :

- To effectively convert text to speech.
- To improve the audio output.
- Improving image processing for better and more accurate results.

e-ISSN: 2582-5208
International Research Journal of Modernization in Engineering Technology and Science
(Peer-Reviewed, Open Access, Fully Refereed International Journal)
Volume:05/Issue:05/May-2023 Impact Factor- 7.868 www.irjmets.com

II. METHODOLOGY

The proposed model consists of two major modules: image processing and audio processing. The image analysis nodes receive an image, either captured in real-time or previously saved in the device, and extract text from it. The extracted text is then processed and converted into high-quality audio output in the voice processing section.The image processing module uses Optical Character Recognition (OCR) to process JPEG or JPG images and generate a text file (.txt) containing the recognized text. OCR utilizes an AI optical mechanism to recognize characters accurately. To improve the accuracy of image recognition, the input image is first converted into binary format before being uploaded to OCR.

The algorithm from UI Path transforms this image, as seen in fig. 2.1.

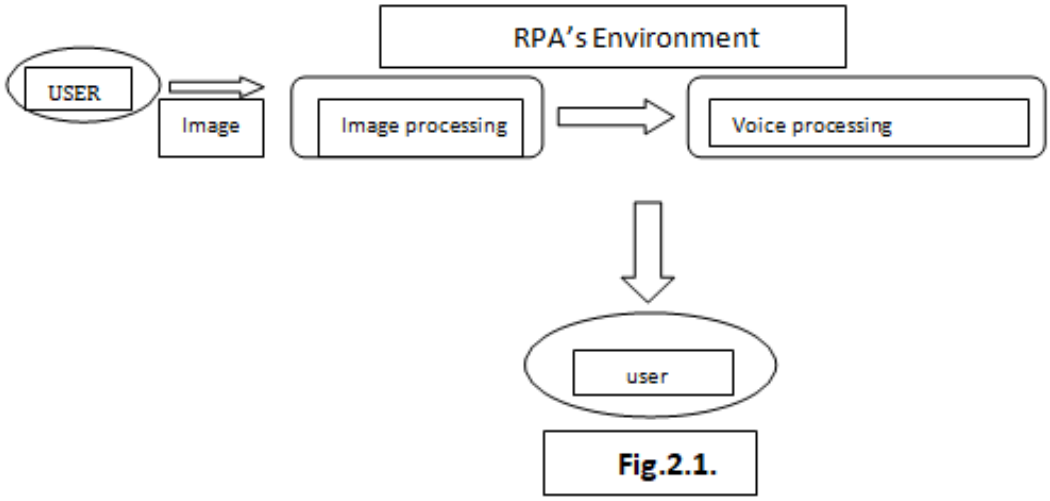


Fig.2.1.

A) IMAGE PROCESSING:-

The pixels on our device's display have different colors, which determine the overall hue of the image. The resolution of the image increases with the number of pixels. To display the final image that we see, our devices arrange these pixels in a specific pattern. As shown in Figure 2, the research is divided into three main steps. The first step is image pre-processing, which is carried out before Optical Character Recognition (OCR) is applied to the image. The OCR algorithm uses the pre- processed image to extract text from it. Lexical analysis is performed to refine the extracted text and eliminate any errors. Once everything is done, the final text file is generated.

fig. 2.2

- STEP 1: Pre-processing of Image:

To facilitate and improve the information extraction process, multiple modifications are made to the received image through preprocessing. The specific techniques utilized for image preprocessing may vary depending on the application.

Binarization:

Converting a color image into a monochrome version is known as binarization. The main goal ofthis process is to highlight the textual information in black and make the background a solid white color, thereby enhancing the accuracy and speed of the OCR system.

Deskewing:

Further processing is required on the binarized image, as the extracted text may not be perfectly aligned. The gathered text is sorted and aligned horizontally, and horizontal alignmentand rotation are applied to correct any misalignment.

e-ISSN: 2582-5208
International Research Journal of Modernization in Engineering Technology and Science
(Peer-Reviewed, Open Access, Fully Refereed International Journal)
Volume:05/Issue:05/May-2023 Impact Factor- 7.868 www.irjmets.com

Despeckle:

It is common for received images to contain noise, which can make it challenging for OCR to accurately recognize text. Therefore, the objective is to enhance text recognition by reducing image noise and improving image quality. This can be achieved by applying image processing techniques such as noise reduction and smoothing.

Line Removal:

To simplify the OCR process, any lines, boxes, and structures that are not text or characters are removed. This eliminates confusion between false lines and characters and improves OCR accuracy, especially when extracting text from tables or documents.

Zoning:

The next step is zoning, where the image is divided into zones representing paragraphs or columns. This helps in determining if the text belongs in a column or the same line, further improving accuracy.

STEP-2: OCR Image Processing:

In the initial step, Optical Character Recognition (OCR) is used to distinguish between character and line spacing. Proper zoning aids in making this process easier, with OCR processing each line or zone of text one at a time. Tokenization involves identifying each empty space between lines of text as a token, which is then treated as a character by the OCR system. Once all the characters have been tokenized, OCR employs two techniques to determine what the characters or tokens represent.

1. Feature Extraction:

The first technique is Feature Extraction, which follows specific guidelines to determine which character each token represents. For instance, a single horizontal line could be 'I', and two equal horizontal lines connected diagonally could represent capital 'N'.

2. Pattern Recognition:

The second technique is Pattern Recognition, in which the previously received tokens are compared to known glyphs. All punctuation marks, digits, alphabets, and special characters have corresponding glyphs, with the best match chosen. The character is then separated into a matrix, and each pixel of the matrix (of the token) is compared with the pixels of the glyphs.

A drawback of this approach is that the OCR system requires the font style and size of the text to be similar to that of the known glyphs. As a result, handwritten text cannot be processed using this method. However, if the typeface of the text is known, this process can be completed more quickly and with greater accuracy.

STEP-3: Image Post-processing:

After the OCR processes the image, a list of words and phrases is generated. The retrieved words are compared against a list of approved terms called a lexicon. Any words that do not match are substituted with the closest match available in the lexicon. This can be useful in distinguishing between visually similar characters like capital 'I' and small 'L', or zero and capital 'O'.

(B) VOICE PROCESSING:

Once the text is retrieved from the image, it is fed into a Text to Speech Synthesizer (TTS) that converts it into audible speech. TTS is a computer-based system that reads out the text using natural pronunciation to produce words and sentences. The generated speech can be heard through speakers or headphones and even saved for future use. This entire process is automated using Robotic Process Automation (RPA) technology. A bot is built to perform all these tasks and provide audio output to the user TTS, or Text to Speech, is a technology that enables computers to read out text in a natural- sounding voice. TTS uses algorithms to analyze the text and create a phonetic representation of the words, which is then converted into speech using a digital voice. The resulting speech can be heard through speakers or headphones and can even be saved for later use.

III. RESULTS AND DISCUSSIONS

Advancements in technology have significantly enhanced user convenience and accessibility. The proposed approach offers a feasible solution to convert text to audio. Users can upload an image directly from their camera or internal storage and let the bot handle the rest. RPA plays a pivotal role in enabling this process, with

e-ISSN: 2582-5208
International Research Journal of Modernization in Engineering Technology and Science
(Peer-Reviewed, Open Access, Fully Refereed International Journal)
Volume:05/Issue:05/May-2023 Impact Factor- 7.868 www.irjmets.com

UI Path providing all the necessary functionalities, as illustrated in Figure 3.1.

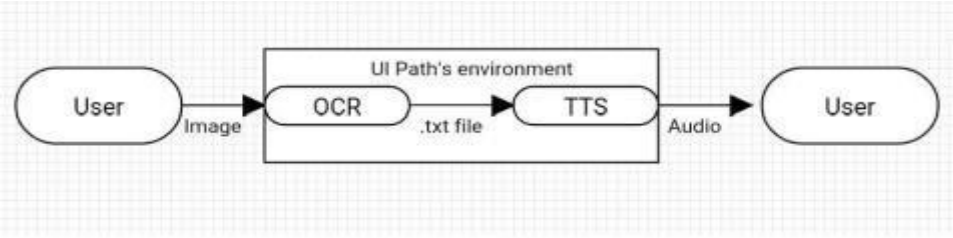


fig 3.1

Once the image is preprocessed, it is then fed into the OCR for text extraction. Beforetext extraction, the image undergoes several processing stages. An example of picturepreparation is shown in Figure 3.2

The first step in picture preparation is binarization. During this step, the image is processed, and a binaryimage is generated, as illustrated in Figure 3.3.

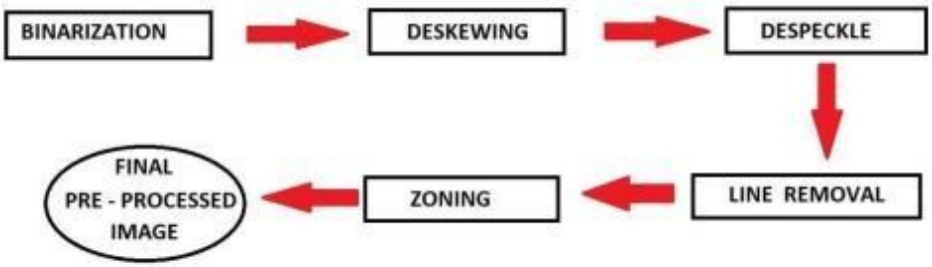
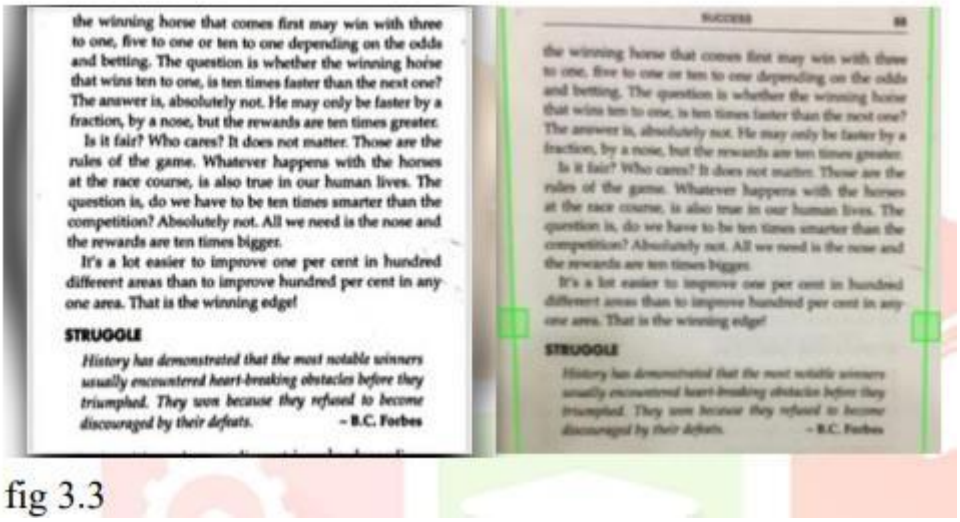


fig. 3.2.

The first step in picture preparation is binarization. During this step, the image is processed, anda binary image is generated, as illustrated in Figure 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

fig 3.3

The technique of width and length is used to straighten the uneven text, as depicted infig.3.4.

e-ISSN: 2582-5208
International Research Journal of Modernization in Engineering Technology and Science
(Peer-Reviewed, Open Access, Fully Refereed International Journal)
Volume:05/Issue:05/May-2023 Impact Factor- 7.868 www.irjmets.com

fig.3.4

Then, the image disturbance is eliminated through a process called despeckle, as shownin fig 3.5

fig 3.5

To ensure accurate OCR output, the unnecessary lines in the image are eliminated, which is referred to as line removal. The image is then divided into zones that represent different columns and paragraphs, as shown in Figure 3.6, after all image processing steps are completed.

e-ISSN: 2582-5208
International Research Journal of Modernization in Engineering Technology and Science
(Peer-Reviewed, Open Access, Fully Refereed International Journal)
Volume:05/Issue:05/May-2023 Impact Factor- 7.868 www.irjmets.com

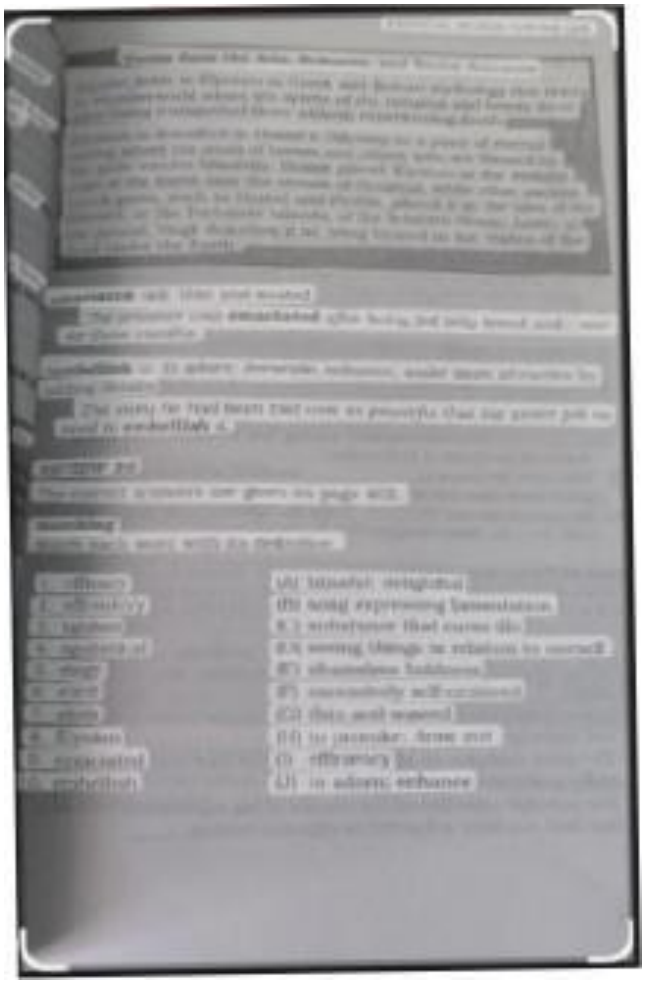


Fig 3.6

Once all image processing is done, the OCR module extracts the text from the processed image and generates a text file. The TTS engine then generates voice outputfrom the text file.

IV. CONCLUSION

The use of RPA in text-to-speech conversion has been demonstrated in this study, with UI Path serving as the platform for OCR image processing and TTS speech processing. This technology can be particularly helpful for individuals who are visually impaired or illiterate. The software's quality and functionality can be enhanced by incorporating multi-language translation capabilities.

V. REFERENCES

[1] Ayushi Trivedi, Navya Pant, Pinal Shah, Simran Sonik and Supriya Agrawal,"Speech to text and text to speech recognition systems" , IOSR Journal of Computer Engineering (IOSR-JCE) e-ISSN: 2278-0661, p-ISSN: 2278-8727, Vol20, Issue2, Ver-I, March. April- 2018.

[2] Hussain Rangoonwala, Vishal Kaushik, P Mohith and Dhanalakshmi Samiappan "Textto speech module International Journal of Pure and Applied Mathematics"Vol115 No. 6 2017.

[3] S. Venkateswarlu1, D. B. K. Kamesh1, J. K. R. Sastry and Radhika Rani , "Text to speech conversion", Indian Journal of Science and Technology", Vol-9(38), October2016.

[4] Gordana Laštovička-Medin, Itana Bubanja, "Hardware approach of text-tospeech in embedded applications" Kaladharan N, An english text to speech conversion system, International Journal of Advanced Research in Computer Science and Software Engineering,, Vol-5, Issue-10, October-2015

[5] Poonam.S.Shetake, S.A.Patil and P.M Jadhav "Review of text to speech conversion methods", International Journal of Industrial Electronics and Electrical Engineering, Vol-2, Issue-8, August-2014

[6] Kaveri Kamble and Ramesh Kagalkar, "Translation of text to speech conversion for hindi language",

"International Journal of Science and Research", Impact Factor (2012): 3.358 Vol- 3 Issue-11, November 2014.

[7] Kumar Patra, Biplab Patra, Puspanjali Mohapatra , "Text to speech conversion with phonematic concatenation," International Journal of Electronics Communication and Computer Technology " (IJECCT) Vol- 2 Issue-5, September- 2012, | Volume 10, Issue 12December 2022 |

[8] Prof. Sheetal A. Nirve,Dr. G. S. Sable,"Optical character recognition for printed text inDevanagari using ANFIS", International Journal of Scientific & Engineering Research, Volume 4, Issue 10, October-2013.

[9] D.Sasirekha, E.Chandra, "TEXT TO SPEECH: A SIMPLE TUTORIAL" International Journal of Soft Computing and Engineering (IJSCE) March 2012.

[10] Priyanka Jose, Govindaru V, "Malayalam Text-to-Speech" International Journal of Engineering and Technical Research (IJETR) Volume-1, Issue-3, May 2013.

International Research Journal Of Modernization in Engineering Technology and Science

(Peer-Reviewed, Open Access, Fully Refereed International Journal)

e-ISSN: 2582-5208

Ref: IRJMETS/Certificate/Volume 05/Issue 05/50500276783

Date: 28/05/2023

Certificate of Publication

This is to certify that author "**Ajay Pal**" with paper ID "**IRJMETS50500276783**" has published a paper entitled "IMAGE2SPEECH RECOGNIZE TEXT IN IMAGES AND CONVERT ACKNOWLEDGED TEXT TO SPEECH" **in International Research Journal Of Modernization In Engineering Technology And Science (IRJMETS), Volume 05, Issue 05, May 2023**

Editor in Chief

IRJMETS Impact Factor 7.868

We Wish For Your Better Future

www.irjmets.com

International Research Journal Of Modernization in Engineering Technology and Science

(Peer-Reviewed, Open Access, Fully Refereed International Journal)

e-ISSN: 2582-5208

Ref: IRJMETS/Certificate/Volume 05/Issue 05/50500276783

Date: 28/05/2023

Certificate of Publication

This is to certify that author "**Vasu Chaudhary**" with paper ID "**IRJMETS50500276783**" has published a paper entitled "IMAGE2SPEECH RECOGNIZE TEXT IN IMAGES AND CONVERT ACKNOWLEDGED TEXT TO SPEECH" **in International Research Journal Of Modernization In Engineering Technology And Science (IRJMETS), Volume 05, Issue 05, May 2023**

Editor in Chief

IRJMETS Impact Factor 7.868

We Wish For Your Better Future

www.irjmets.com

International Research Journal Of Modernization in Engineering Technology and Science

(Peer-Reviewed, Open Access, Fully Refereed International Journal)

e-ISSN: 2582-5208

Ref: IRJMETS/Certificate/Volume 05/Issue 05/50500276783

Date: 28/05/2023

Certificate of Publication

This is to certify that author "**Anmol Tyagi**" with paper ID "**IRJMETS50500276783**" has published a paper entitled "IMAGE2SPEECH RECOGNIZE TEXT IN IMAGES AND CONVERT ACKNOWLEDGED TEXT TO SPEECH" **in International Research Journal Of Modernization In Engineering Technology And Science (IRJMETS), Volume 05, Issue 05, May 2023**

Editor in Chief

IRJMETS Impact Factor 7.868

We Wish For Your Better Future

www.irjmets.com

International Research Journal Of Modernization in Engineering Technology and Science

(Peer-Reviewed, Open Access, Fully Refereed International Journal)

e-ISSN: 2582-5208

Ref: IRJMETS/Certificate/Volume 05/Issue 05/50500276783

Date: 28/05/2023

Certificate of Publication

*This is to certify that author "**Rahul Singh**" with paper ID "**IRJMETS50500276783**" has published a paper entitled "IMAGE2SPEECH RECOGNIZE TEXT IN IMAGES AND CONVERT ACKNOWLEDGED TEXT TO SPEECH" **in International Research Journal Of Modernization In Engineering Technology And Science (IRJMETS), Volume 05, Issue 05, May 2023***

Editor in Chief

IRJMETS Impact Factor 7.868

We Wish For Your Better Future

www.irjmets.com