TEXT READER FOR VISUALLY IMPAIRED PEOPLE

(Saleha Saudagar, Komal Bhat, Sarvesh Koli, Piyush Kothekar, Prajwal Korade , Ayushi Kowe)
Vishwakarma Institute of Technology, Pune, India

(saleha.saudagar@vit.edu, sarvesh.koli21@vit.edu, piyush.kothekar21@vit.edu, prajwal.korade21@vit.edu, komal.bhat21@vit.edu, ayushi.kowe21@vit.edu)

***Abstract*—This paper presents camera based system which will help blind person for reading text patterns printed on newspapers ,blogs, notice board . The detected text is compared with the template and converted into the speech output. The text patterns are localized and binarized using Optical Character Recognition (OCR). The recognized text is converted to an audio output. The speech output is given to the blind user. Therefore, this paper deals with analysis of detection and recognition of different text patterns on different objects.**

***Keywords:*-Pyttessract,Opencv,Optical Character Recognition (OCR),Text to speech conversion**

I. INTRODUCTION

Nowadays Text-to-speech (TTS) technology has greatly improved the accessibility of written information for blind and visually impaired individuals. With TTS, written text is converted into audio, allowing blind users to listen to books, articles, and other written materials.

Before the development of TTS, blind individuals had to rely on braille or human readers to access written information. While these methods are still important and widely used, TTS has provided a new level of independence and convenience for blind individuals. With TTS software, users can easily access and interact with written materials on their own, without the need for a sighted individual to assist them. In addition to providing access to written materials, TTS technology has also opened up new opportunities for education and employment for blind individuals. Text-to-speech (TTS) technology has greatly improved the accessibility of written information for blind and visually impaired individuals. With TTS, written text is converted into audio, allowing blind users to listen to books, articles, and other written materials. However, TTS technology has limitations, as it requires written text to be input into the system. This can be a barrier for blind individuals who need to access information that is only available in an image format, such as a textbook or a scanned document. To overcome this barrier, OCR (optical character recognition) technology can be used to extract the written text from an image and convert it into a format that can be processed by a TTS system. By combining OCR and TTS technology, it is possible to create a reader for blind individuals that can convert images to speech, greatly improving their accessibility to written information. Using an image-to-speech reader has numerous benefits for blind individuals.

It allows them to access a wider range of written materials, including books, articles, and documents that may not be available in braille or audio formats. It also provides them with more independence and convenience, as they can access and interact with written materials on their own without the need for a sighted individual to assist them. Overall, an image-to-speech reader for blind individuals can greatly enhance their access to written information and improve their independence and convenience in accessing and interacting with written materials.

II. LITERATURE REVIEW

[1] Real Time Text Detection and Recognition on Hand Held Objects to Assist Blind People.

Held Objects to Assist Blind PeopleSamruddhi Deshpande, Ms. Revati Shriram

This prototype system reads text from the image that was captured by the camera. Three functional sections make up the framework: scene capture, data processing, and audio output. the setting capture elements gather pictures of various things and extract the object of interest from the picture.

A camera is employed in this system to capture scenes. The data processing component is used to extract text information from images by detecting text patterns. To extract text information from an image, MSER is employed. By doing this, the text patterns on various backdrops will be extracted. This is the text's reliable algorithm.

detection. On the localized text patterns, OCR is employed for text recognition. For blind users, the system will translate the localized text to audio output. The results show that employing MSER and OCR for text detection and recognition, great performance may be achieved.

[2] AI based Reading System for Blind using OCR

In this paper, this program uses the camera to recognize the text, scans it, and then turns it into digital text that the system can read and display. It provides speech output and translates text. A fundamental understanding of AI and OCR is necessary in order to comprehend the project's dynamics. This report shows how Language Translator functions in its entirety and outlines the system prerequisites.

[3] Character Detection and Recognition System for Visually Impaired People

This paper describes an approach to extract and recognize text from scene images effectively using computer vision technology and to convert recognized text into speech so that it can be incorporated with hardware to develop Electronic travel aid for visually impaired people in future.

[4] Image to Multilingual Text Conversion for Literacy Education

This paper examines the widely used methods for translating images into text. by filling in the gaps that a careful assessment of the literature revealed. The purpose of this software is to aid travelers worldwide. It extracts and recognizes images using the OCR method. It will first capture an image, extract the text from it, and then identify the characters. When this cycle of processes is finished, it searches its database for the text that needs to be translated. If the text is found in the database, it is translated; otherwise, the application will automatically look for the content online.

[5] Visual Assistance for Blind using Image Processing

This research describes a cutting-edge method for helping persons who are blind. The suggested system's straightforward architecture and user-friendliness make the topic more accessible independently minded at home. The technology also seeks to assist the blind in navigating their environment by spotting obstacles, finding their essentials, and reading signs and texts. Initial tests have produced encouraging results, allowing the user to safely and freely move about his environment. By allowing speech to be used as the input to access his basic needs, the system is made much more user-friendly.

[6] Comparative Analysis of Text Extraction from Color Images using Tesseract and OpenCV

In this paper ,They have tried to apply various image processing techniques. One of them Tessesrect method will allow us to recognize text from most types of background. We propose to provide methods for easy text extraction.

[7] A Review On Conversion Of Image To Text As Well As Speech Using Edge Detection And Image Segmentation

Core idea for image to text and speech conversion is to overcome the challenges faced by a blind person in real life. The techniques of image segmentation and edge detection play an important role in implementing this system. We formulate the interaction between image segmentation and object recognition in the framework of Canny algorithm. The system goes through various phases such as preprocessing, feature extraction, object recognition, edge detection, image segmentation and text-to-speech (TTS) conversion. The database of this system consists of huge set of sample images which help to identify similar kind of objects in every different image. The system mainly consists of two main modules such as image-to-text and text-to-speech. An image-to-text module generates text descriptions in natural language based on understanding of image. A text-to-speech module converts natural language into speech synthesis.

[8] Text Extraction from Image and Text to Speech Conversion

The majority of the photos have text regions that can be successfully detected by the suggested method, and the text can be accurately extracted from the detected regions. According to experimental research, the suggested method can accurately identify text sections from photos with a variety of font sizes, styles, and colors. Even while the method circumvents most of the problems that other algorithms go into, it still struggles to work on photos with small or blurry text sections.

The word-confidences for the words they were able to obtain after using optical character recognition on the image used in this paper's experimental study are shown.

[9]Text to Speech Conversion using Optical Character Recognition introduced by Shtavani S., Usha A., Karthik P., Mohan Babu C.

In this paper authors have splitted their work in three components which are camera which will work as an eye to their system a programmable system which will help in programming the system in the desired manner and the last one is to show the output collected. Authors have proposed an OCR system which consists of Rasberry pi which is a compact sized as well as low weight system which will help in creating an OCR system. Here the RB pi is connected to programmable system in this case which is computer by an ethernet cable. The camera used for image capturing can be RB pi camera or it can be web cam as well. Here authors have used the web cam for image capture. This was about the components used and for cherry on top the software systems used here are first one is OCR used to extract the text from image provided to camera and the TTS text to speech which is used to read out text aloud. Here authors have used pytesseract to do so. To explain about methodology authors have done Image 1.Pre-Processing which consist of few steps which are Binarization, Deskew, Despeckle, Line removal, Zoning . Again authors have also done2. Processing the image In which they have a=included Pattern recognition and feature extraction of image. And at the end 3. Post processing the image Lexical Restriction have been done .

All process is done by storing content extracted from images in files. At first output from OCT is stored in file names speech.txt which is text file and then the file us converted to voice module.

III. TECHNOLOGIES USED

Open CV

OpenCV (Open Source Computer Vision) is a free and open-source library of computer vision and machine learning algorithms.

It can be used to develop applications that can process images and videos to identify and track objects, recognize faces, and perform other image processing tasks.

To use OpenCV to create a text reader for blind individuals, you would first need to capture an image of the text using a camera or scanner.

Then, you could use OpenCV to preprocess the image, such as by converting it to grayscale and applying image enhancement techniques to improve the contrast and clarity of the text.

Next, you could use OpenCV's OCR (optical character recognition) module to extract the text from the image.

This module uses machine learning algorithms to recognize and classify the individual characters in the image.

Finally, you could use a TTS (text-to-speech) engine to convert the extracted text into an audio file, which could then be played back to the user through headphones or a speaker.

Using OpenCV to create a text reader for blind individuals would allow you to build a low-cost and user-friendly solution for accessing written materials.

It would also provide the user with the ability to customize the reading experience, such as by adjusting the speed and pitch of the TTS voice.

Pytesseract

Pytesseract is a Python wrapper for Tesseract OCR (Optical Character Recognition), an open-source library for extracting text from images.

Pytesseract allows you to use Tesseract OCR from within your Python scripts, making it easy to integrate Tesseract OCR into your workflow.

Tesseract OCR is a highly accurate and reliable OCR engine, with support for over 100 languages and a wide range of image types and formats.

It has been developed and maintained by Google since 2006, and is widely used in a variety of applications including document scanning, digital archiving, and document processing.

Pytesseract provides a number of additional options and functions that allow you to customize and fine-tune the OCR process.

For example, you can specify the language of the text, the page segmentation mode, or the whitelist/blacklist of characters to include/exclude.

You can also use the pytesseract.image_to_data() function to extract additional information such as the bounding boxes and confidence scores of the recognized characters.

Optical Character Recognition (OCR)

Optical Character Recognition (OCR) is a technology that allows computers to extract text from images and documents.

OCR has a wide range of applications, including document scanning, digital archiving, document

processing, and accessibility for the visually impaired.

OCR works by analyzing the pixels in an image or document and identifying patterns and features that correspond to specific characters or symbols.

The OCR software then converts these patterns into digital text that can be edited, searched, and processed by computers.

There are many different OCR software and libraries available, ranging from open-source to commercial. Some of the most popular OCR software and libraries include Tesseract OCR, GOCR, Ocrad, and CuneiForm.

OCR accuracy depends on a number of factors, including the quality and clarity of the input image or document, the font and layout of the text, the language of the text, and the capabilities of the OCR software.

OCR accuracy can be improved by preprocessing the input image or document to enhance contrast, remove noise, and correct distortion, as well as by adjusting the OCR parameters and settings to optimize for the specific characteristics of the text.

OCR technology has come a long way in recent years, but there is still room for improvement in terms of accuracy and robustness.

Researchers are working on developing new OCR algorithms and techniques that can better handle difficult images such as those with low contrast, noise, or distorted text, as well as languages with complex scripts.

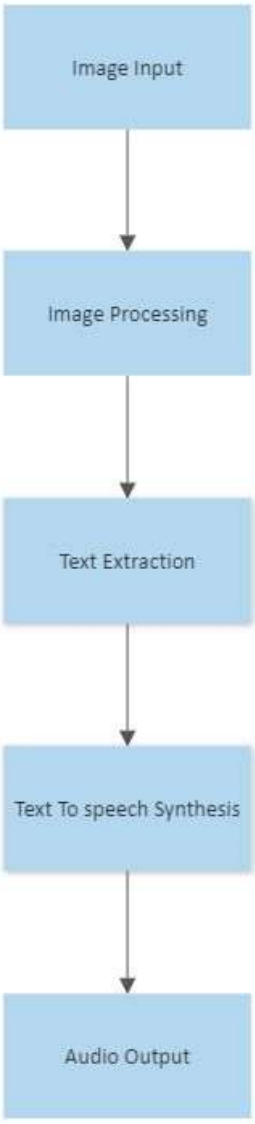

Fig. 1. Flow Chart

IV. METHODOLOGY AND EXPERIMENTATION

As shown in Fig.1, the proposed system is divided into 4 stages:-

1.Image Acquisition
2.Image Processing
3.Text Detection and Recognition
4.Text to speech conversion

1.Image Acquisition:

Image acquition involves retrieving an image from a source, usually hardware systems like cameras, sensors, etc

In this paper, with the help of opencv library in python ,we are taking real time screenshot from the live web camera .The captured screenshot is saved and stored in local folder with the help of Os module in python. Each and every time ,We are retrieving the recently captured screenshot from

that stored path in folder and taking it as a input for further processing .

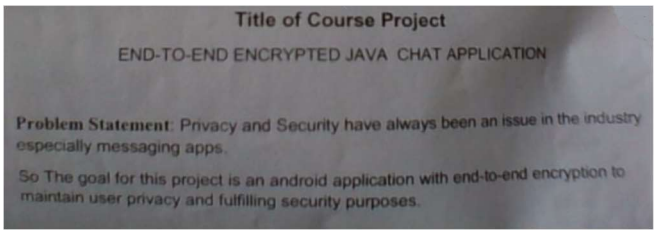

Fig.2.Taken Screenshot before Text Recognition

2.Image Processing: The input image is preprocessed to improve the quality and contrast of the text, it also involves clearing the background of the image which can make it easier for the algorithm to recognize the each character or word . This involve techniques such as Text localization, image resizing, ,image compression, image manipulation, image generation,

3.Text Detection and Recognition: The OCR algorithm analyzes the input image to locate and identify areas of the image that contain text. This may involve techniques such as edge detection and connected component analysis .Once the text has been detected, the algorithm processes the text to recognize and extract the individual characters with the help of pytteseract library. This may involve techniques such as text background ,pattern matching, text formatting, mode of recognition

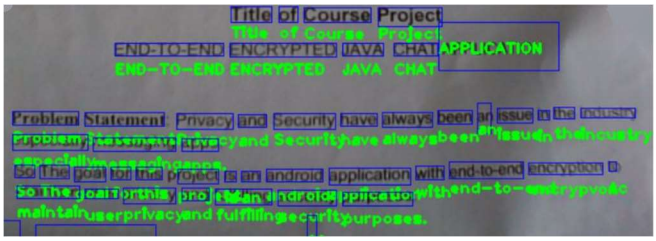

Fig.3. Printed text image after Text Recognition

4.Text to speech Conversion:
Text to speech involves conversion of detected text into speech . With the help of the pyttsx3 library in python ,the recognized text is converted into speech as audio which given to the visually impaired people.

V. FUTURE SCOPE

1.Improved OCR accuracy:
Researchers are working on developing OCR algorithms and techniques that can better handle a wider range of image types and conditions, including low resolution, noise, blur, distortion, and complex layout.
They are also working on improving the accuracy of OCR for languages with complex scripts, such as Arabic, Chinese, and Devanagari.
2.Enhanced text-to-speech quality:
Researchers are working on developing text-to-speech algorithms and techniques that can produce more natural and expressive speech, including variations in prosody, intonation, and accent.
They are also exploring ways to improve the quality of text-to-speech synthesis for languages with complex phonology and prosody, such as tonal languages and languages with vowel harmony.
3.Integration with machine translation:
Image-to-speech conversion can be integrated with machine translation to create systems that can automatically translate foreign language text into speech.
This could be useful for travelers, students, or anyone else who needs to understand written materials in a different language.
4.Integration with language processing:
Image-to-speech conversion can be integrated with language processing technologies such as part-of-speech tagging, named entity recognition, and syntactic parsing to create more sophisticated and context-aware speech synthesis systems.

VI. CONCLUSION

In conclusion, image-to-speech conversion is a useful and valuable technology that has many potential applications and benefits. It can be used to make printed materials more accessible for the visually impaired, to translate foreign language text into speech, to transcribe spoken words into text, and to create audio versions of educational materials, e-books, and other content. While there has been significant progress in the field of image-to-speech conversion in recent years, there is still room for improvement in terms of accuracy and quality. Researchers are working on developing new OCR (optical character recognition) algorithms and techniques that can better handle difficult images and languages, and text-to-speech

algorithms and techniques that can produce more natural and expressive speech. There are also many potential areas for future research and development in the field of image-to-speech conversion, including the integration of image-to-speech with other technologies such as machine translation and language processing, the development of more sophisticated and context-aware speech synthesis systems, and the exploration of new applications and services. Overall, the potential of image-to-speech conversion is vast, and it has the potential to transform the way we interact with and access information. As research and development in this field continues to advance, we can expect to see even more exciting and innovative applications of image-to-speech conversion in the future.

VII. REFERENCES

[1] S. Deshpande and R. Shriram, "Real time text detection and recognition on hand held objects to assist blind people," 2016 International Conference on Automatic Control and Dynamic Optimization Techniques (ICACDOT), 2016, pp. 1020-1024, doi: 10.1109/ICACDOT.2016.7877741.

[2] A. Mathur, A. Pathare, P. Sharma and S. Oak, "AI based Reading System for Blind using OCR," 2019 3rd International conference on Electronics, Communication and Aerospace Technology (ICECA), 2019, pp. 39-42, doi: 10.1109/ICECA.2019.8822226.

[3] Panchal, Akhilesh & Varde, Shrugal & Panse, Meena. (2016). Character detection and recognition system for visually impaired people. 1492-1496. 10.1109/RTEICT.2016.7808080.

[4] M. Ajmal, F. Ahmad, M. -E. A.M., M. Naseer, A. Muhammad and M. Ashraf, "Image to Multilingual Text Conversion for Literacy Education," 2018 17th IEEE International Conference on Machine Learning and Applications (ICMLA), 2018, pp. 1328-1332, doi: 10.1109/ICMLA.2018.00215.

[5] B. Deepthi Jain, S. M. Thakur and K. V. Suresh, "Visual Assistance for Blind Using Image Processing," 2018 International Conference on Communication and Signal Processing (ICCSP), 2018, pp. 0499-0503, doi: 10.1109/ICCSP.2018.8524251.

[6] A. Revathi and N. A. Modi, "Comparative Analysis of Text Extraction from Color Images using Tesseract and OpenCV," 2021 8th International Conference on Computing for Sustainable Global Development (INDIACom), 2021, pp. 931-936.

[7] Kagalkar, Ramesh. (2014). A Review On Conversion Of Image To Text As Well As Speech Using Edge Detection And Image Segmentation.

[8] https://www.ijert.org/research/text-extraction-from-image-and-text-to-speech-conversion-IJERTCONV9IS03002.pdf

[9] https://ijcrt.org/papers/IJCRT2108410.pdf

[10]https://www.irjet.net/archives/V7/i7/IRJET-V7I7994.pdf