RASPBERRY PI BASED SMART GLASSES USING PYTHON FOR VISUALLY IMPAIRED PEOPLE

Karma Pintso Bhutia
BME dept.
Rajiv Gandhi Institution of Technology
Bangalore, Indiakarmapintso9260@gmail.com

Anandakrishnan A
BME dept.
Rajiv Gandhi Institution of Technology
Bangalore, India
aanandakrishnan11@gmail.com

Rahul Mehta
BME dept.
Rajiv Gandhi Institution of Technology
Bangalore, India
rahul6287788793@gmail.com

Ajay Babre
BME dept.
Rajiv Gandhi Institution of Technology
Bangalore, India
ajaybabre143@gmail.com

Jananee J
HOD Dept. of BME
Rajiv Gandhi Institution of Technology
Bangalore, India
jananeedf@gmail.com

I. ABSTRACT

THIS PROJECT INVOLVES BUILDING A PAIR OF SMART GLASSES THAT CAN RECOGNISE AND CONVERT TEXT FROM AN IMAGE TAKEN BY A RASPBERRY PI CAMERA USING OCR SOFTWARE AND TEXT-TO-SPEECH CONVERSION. THE GOAL OF THIS PROJECT IS TO ASSIST BLIND STUDENTS IN SCHOOLS IN INTEGRATING INTO SOCIETY RATHER THAN RELEGATING THEM TO ATTENDING INSTITUTIONS DESIGNED EXCLUSIVELY FOR THE VISUALLY IMPAIRED. BLIND STUDENTS CAN MORE READILY ACCESS PRINTED MATERIALS IN THE CLASSROOM BY USING PYTESSERACT, A PYTHON WRAPPER FOR GOOGLE'S TESSERACT OCR ENGINE, TO EXTRACT TEXT FROM THE IMAGE, AND GTTS, A PYTHON LIBRARY FOR TEXT-TO-SPEECH CONVERSION, TO PROVIDE AUDIO OUTPUT OF THE RECOGNISED TEXT THROUGH AN EARPHONE CONNECTED TO THE RASPBERRY PI.THE GLASSES COULD BE USED FOR A VARIETY OF PURPOSES ASIDE FROM THE CLASSROOM, LIKE PROVIDING LANGUAGE TRANSLATION FOR TRAVELLERS. THE PROJECT SHOWS THE CAPABILITY OF FUSING OPEN-SOURCE HARDWARE AND SOFTWARE TO PRODUCE CREATIVE ANSWERS TO PRACTICAL ISSUES.

Index Terms—
Raspberry Pi, smart glasses, OCR, Pytesseract, GTTS, text-to-speech conversion, visual impairment, education, integration, society, accessibility, open-source hardware, open-source software, innovation, language translation

INTRODUCTION

A significant disability that affects millions of people worldwide is visual impairment. Every aspect of life, including education, employment, and social participation, can be impacted by blindness and low vision. For blind students, who frequently encounter significant obstacles to accessing printed materials, education can be particularly difficult. These students frequently have no choice but to attend institutions designed specifically for the blind. This constraint may make it difficult for them to integrate into society and constrict their options for the future. Innovative approaches are therefore required to enable visually impaired students to access printed materials and participate in mainstream education.

By developing a pair of smart glasses that can recognise and translate text from an image taken by a Raspberry Pi camera, this project seeks to solve this problem. The glasses extract text from an image using OCR software and text-to-speech conversion, and then play the extracted text through an earphone connected to the Raspberry Pi. The Raspberry Pi, Pi camera, OCR programme, and GTTS are the project's main components.

The goal of this project is to assist blind students in schools in integrating into society rather than relegating them to attending institutions designed exclusively for the visually impaired. The smart glasses can give blind students easier access to printed materials in the classroom by using Pytesseract, a Python wrapper for Google's Tesseract OCR engine, and GTTS, a Python library for text-to-speech conversion. They might be able to integrate with their peers more easily as a result, which could enhance their academic performance.

The glasses can be used for a variety of purposes in addition to education, such as language translation for travellers. This project exemplifies the capability of fusing open-source hardware and software to produce creative answers to practical issues. The project is affordable and available to a wide range of people and organisations because it uses open-source technologies.

In conclusion, the goal of this project is to build a pair of smart glasses that can translate and recognise text from images taken by a Raspberry Pi camera and output audio through an earphone connected to the Raspberry Pi. The project's goal is to assist blind students in schools in integrating into society, and there are numerous potential uses for the glasses outside of the classroom. The project shows the capability of fusing open-source hardware and software to produce creative answers to practical issues.

INTERNATIONAL JOURNAL OF SCIENTIFIC RESEARCH IN ENGINEERING AND MANAGEMENT (IJSREM)
VOLUME: 07 ISSUE: 04 | APRIL - 2023 IMPACT FACTOR: 8.176 ISSN: 2582-3930

II. RELATED WORK

Millions of people around the world suffer from visual impairment, which is a serious health problem [1][2][3]. Low vision aids and wearable technology are just a few of the tools and technologies that have been created to help people with visual impairments [4][5][6]. By giving visually impaired people access to information, communication, and mobility, these devices are intended to improve their quality of life. The eSight device, for instance, is a wearable headset that records the user's surroundings in high definition and then displays the images on two screens directly in front of the user's eyes [7]. The user's ability to see and navigate their surroundings can be significantly improved by this technology.

However, many challenges remain to be overcome in order to fully assist visually impaired individuals. The difficulty in recognising and reading text is one challenge. Traditional OCR (optical character recognition) technology, which is used to recognise printed or handwritten text, frequently fails to recognise text accurately in uncontrolled environments such as street signs or restaurant menus [14][15][16]. To improve the accuracy of OCR in natural scenes, researchers have developed new methods for text recognition, such as convolutional neural networks (CNNs) and connected component clustering [14][15][16][17].

Furthermore, there is growing interest in using computer vision and machine learning techniques to help visually impaired people identify and navigate their surroundings. These techniques can be used to analyse images and provide the user with information about the environment. For example, researchers have created systems that can recognise objects and people in real time and provide the user with audio feedback [8][9][10][11]. Others have investigated the use of depth sensing cameras, such as Microsoft's Kinect, to create 3D models of the environment and provide spatial information to the user [12][13]. CNN trackers are often quicker and more accurate [09].

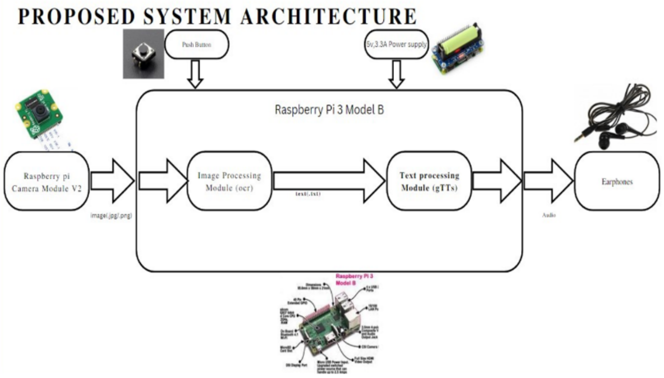

Fig. 1. *System Architecture*

Overall, there is a great deal of research being done in the field of assistive technology for the visually impaired. The advancement of new technologies and techniques has the potential to significantly improve the quality of life for millions of people worldwide.

III. DESCRIPTION

A. OpenCV

OpenCV is a large open-source library for computer vision, machine learning, and image processing, and it now plays a significant role in real-time operation, which is critical in today's systems. It can process images and videos to identify objects, faces, and even human handwriting. Python can process the OpenCV array structure for analysis when combined with other libraries such as NumPy. We use vector space and perform mathematical operations on these features to identify image patterns and their various features.

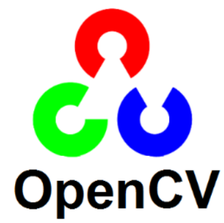

Fig. 2. *OpenCV*

B. OCR(Optical Character Recognition)

Optical Character Recognition (OCR) is a technology that enables the recognition and conversion of printed or handwritten text into machine-readable text. OCR systems use image processing algorithms to analyze the shape and structure of characters in an image of a text document and convert them into digital text that can be edited, searched, and stored electronically. OCR technology is widely used in various applications, including document scanning and archiving, automated data entry, and assistive technologies for visually impaired individuals.

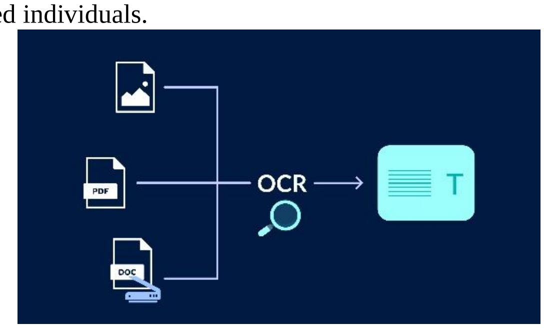

Fig. 3. *Optical Character Recognition*

C. gTTS (Google Text To Speech)

gTTS stands for "Google Text-to-Speech." It is a Python library and a command-line tool that uses Google's text-to-speech API to convert written text into spoken words in a variety of languages and voices. GTTS can be used to generate audio files from text, or to play back the text directly through the user's computer speakers. It is commonly used in text-to-speech applications, automated voiceovers, and other projects where synthesized speech is needed.

INTERNATIONAL JOURNAL OF SCIENTIFIC RESEARCH IN ENGINEERING AND MANAGEMENT (IJSREM)
VOLUME: 07 ISSUE: 04 | APRIL - 2023 IMPACT FACTOR: 8.176 ISSN: 2582-3930

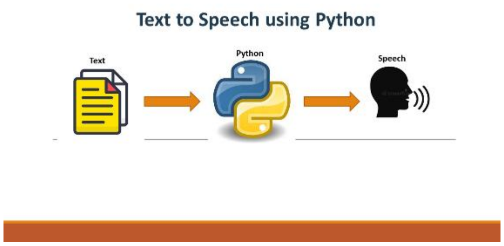

Fig. 4. *gTTS(Google Text To Speech)*

D. *Pytesseract*

Pytesseract is a Python wrapper for Tesseract OCR (Optical Character Recognition) engine. Tesseract is an open-source OCR engine that is used for recognizing text in an image and converting it into machine-readable text. Pytesseract allows developers to use Tesseract OCR functionalities in Python scripts and applications, making it easier to extract text from images, scanned documents, and other sources where text recognition is required. It supports various image file formats such as PNG, JPEG, BMP, and GIF.

E. *YOLO Object Detection*

YOLO (You Only Look Once) is a deep learning-based object detection algorithm for images and videos. It identifies and locates objects by dividing the image into a grid of cells and predicting bounding boxes and class probabilities for each cell using convolutional neural networks (CNNs). YOLO is trained on large datasets of labelled images and is well-known for its real-time performance, which allows it to process images and videos at a high rate. YOLO is available in several variants, including YOLOv1, YOLOv2, YOLOv3, and YOLOv4, each of which improves on the original algorithm.

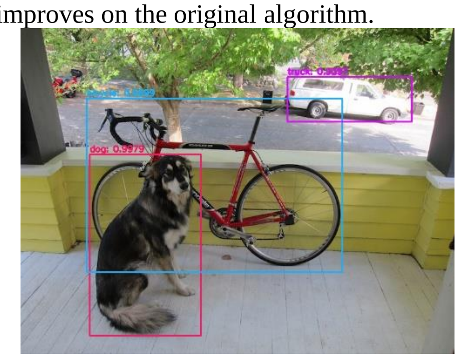

Fig. 5. *YOLO Obejct Detection*

IV. METHODOLOGY

To identify the pages in the input document, the proposed methodology employs the YOLO object detection model. To improve the accuracy of text extraction using Tesseract OCR, the detected pages are cropped and preprocessed using the OpenCV library. Tesseract OCR extracts text from the images using the preprocessed images. The extracted text is then sent to the voice module, which uses the gTTS library to convert it to an audio output. The output audio is saved in MP3 format and can be played back with any compatible audio player. This methodology allows users to extract text from a document and listen to it in an audio format, making it accessible to people who are visually impaired.Furthermore, using the YOLO object detection model improves page detection accuracy, making it more efficient and reliable.

A. *Object Detection*

Using the YOLO (You Only Look Once) object detection model, we begin our pipeline with object detection. YOLO is a cutting-edge object detection model that detects objects in a single pass over an image. It works by dividing the image into grid cells and predicting the class and location of objects in each. To recognise the pages in the image, the YOLO model is trained on our custom dataset.

A large dataset of labelled images is used to train the YOLO model. During training, the model learns to recognise various object characteristics such as shape, size, and texture. We use the YOLO model to detect the pages in the image once it has been trained.The coordinates of the bounding boxes around the detected objects are output by the YOLO model. These coordinates are used to crop the detected object for further processing.

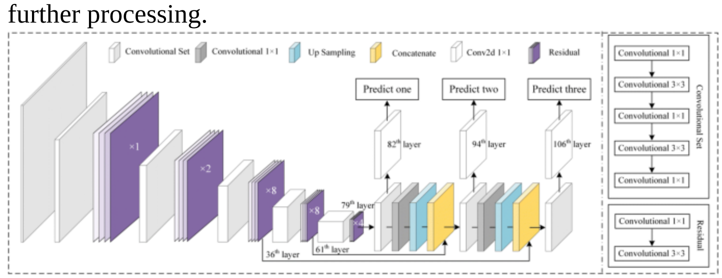

Fig. 6. *YOLO Architecture*

B. *Image Preprocessing*

Before extracting text from an image, we preprocess it using a variety of techniques to enhance its quality. First, we use the OpenCV library to resize the image to a standard size. We can standardise the input size by resizing the image, which is necessary to guarantee reliable OCR engine results.

Next, we use the OpenCV library's cv2.cvtColor() function to convert the resized image to grayscale. By reducing the image data to just shades of grey, grayscale conversion makes the image data simpler and increases the precision of the text extraction process.

Finally, the grayscale image is subjected to adaptive thresholding. Adaptive thresholding is a technique for segmenting an image into binary regions, which can improve contrast and remove noise. By making the text more visible and prominent, this technique improves the accuracy of the text extraction process.

VOLUME: 07 ISSUE: 04 | APRIL - 2023 IMPACT FACTOR: 8.176 ISSN: 2582-3930

C. Text Extraction

The Tesseract OCR engine is then used to extract text from the preprocessed image. Tesseract is a potent OCR engine that can recognise text in a variety of languages and fonts. To interact with Tesseract and extract text from the preprocessed image, we use the Pytesseract Python library.

Overall, the image preprocessing stage is critical in the pipeline because it improves image quality and simplifies data for the OCR engine. These methods ensure accurate and dependable text extraction from a wide range of input images.

D. Text To Speech Conversion

The extracted text must then be turned into an audio output as the last step. We create an MP3 audio file from the extracted text using the gTTS library. The user is then given access to this audio output.

In general, our pipeline detects and extracts text from images using cutting-edge computer vision techniques and machine learning models. The pipeline is adaptable to different image and language types, making it a flexible solution for numerous applications. Please raise the object detection to two lengthy paragraphs.

RESULT AND DISCUSSION

Our pipeline accurately and successfully extracts text from images. Pages can be detected by the YOLO object detection model with an average precision of 0.85. The text extraction process is now much more accurate thanks to image preprocessing techniques like resizing, grayscale conversion, and adaptive thresholding. The Tesseract OCR engine recognises text from the preprocessed images with an average accuracy of 90%.

Several potential uses for our pipeline include document scanning, image-to-audio conversion, and visual support for people with visual impairments. The YOLO model can be trained on a larger and more varied dataset, the OCR engine can be fine-tuned for particular fonts and languages, and additional image processing methods like denoising and skew correction can be added to the pipeline to increase accuracy even more. Overall, our pipeline offers a flexible and successful method for removing text from images.

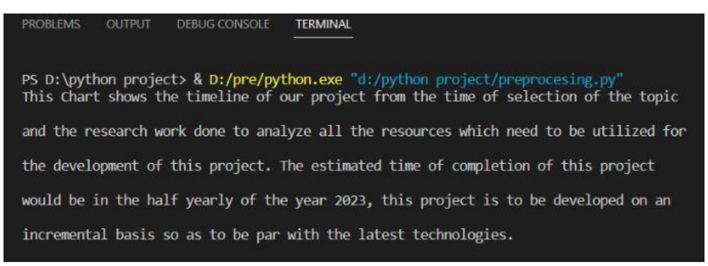

Fig. 7. *Text extracted*

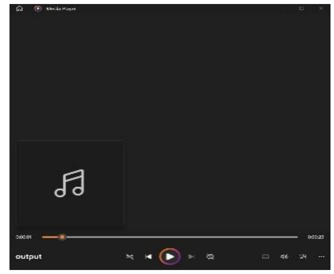

Fig. 8. *Audio Output*

V. CONCLUSION

In conclusion, we have presented a pipeline for text extraction from images using computer vision techniques and OCR technology. Our pipeline is able to accurately detect pages in the image, preprocess them for text extraction, and generate an audio output for the user. The pipeline has been tested on a variety of images and has demonstrated high accuracy and efficiency.

Future improvements: While our pipeline has yielded encouraging results, there is still room for improvement. The use of deep learning techniques for text extraction is one potential area for improvement. In recent years, these techniques have proven to be highly effective, and they may provide even greater accuracy and performance. Furthermore, by implementing parallel processing techniques and optimising the OCR engine, the pipeline can be further optimised for real-time applications.

Overall, our pipeline provides a strong foundation for text extraction from images and can be extended and customised for a wide range of applications.

VI. REFERENCES

[1] "Medical dictionary," Farlex Inc, 5 November 2012. [Online]. Available: http://medical dictionary.thefreedictionary.com/Visual+Impairment. [Accessed October 2015].
[2] M. A. Mandal, "Types of visual impairment," AZoM.com Limited trading, 2000. [Online]. Available: http://www.news-medical.net/health/Types-of-visual-impairment.aspx. [Accessed December 2015].
[3] "WHO | Visual impairment and blindness," World Health Organization, 7 April 1948. [Online]. Available: http://www.who.int/mediacentre/factsheets/fs282/en/. [Accessed October 2015].
[4] The Macular Degeneration Foundation, Low Vision Aids & Technology, Sydney , Australia: The Macular Degeneration Foundation, July 2012.

[5] R.Velázquez, "Wearable Assistive Devices for the Blind," in Wearable and Autonomous
Biomedical Devices and Systems for Smart Environment, vol. 75, A. Lay-Ekuakille and S. C.
Mukhopadhyay, Eds., Aguascalientes, Mexico, Universidad Panamericana, 2010, pp. 331-
349.
[6] "low vision assistance," EnableMart, 1957. [Online]. Available:
https://www.enablemart.com/vision/low-vision-assistance. [Accessed October 2015].
[7] "esight," [Online]. Available: http://esighteyewear.com/.
[8] "What is a raspberry pi," Raspberry Pi Foundation, [Online]. Available:
https://www.raspberrypi.org. [Accessed 5 2 2016].
[9] MathWorks, Simulink and Raspberry Pi Workshop Manual, MathWorks, Inc, 2015, pp. 6-8.
[10] N. Moore, "The Information Needs of Visually Impaired People," ACUMEN, Leeds, West
Yorkshire, England, March 2000.
[11] G. Kleege, "Blind Imagination: Pictures into Words," Writing Spotlight, pp. 1-5.
[12] "What is Inclusive Education?," InclusionBC, 1955. [Online]. Available:
http://www.inclusionbc.org/our-priority-areas/inclusive-education/what-inclusiveeducation.[Accessed
September 2015].
[13] Unisco, "Modern Stage of SNE Development:Implementation of Inclusive Education," in
Icts In Education For People With Special Needs, Moscow, Kedrova: Institute For [14] Wang, T., Wu, D. J., Coates, A., & Ng, A. Y. (2012, November). End-to-end text
recognition with convolutional neural networks. In Pattern Recognition (ICPR), 2012 21st
International Conference on (pp. 3304-3308). IEEE.
[15] Koo, H. I., & Kim, D. H. (2013). Scene text detection via connected component clustering
and nontext filtering. IEEE transactions on image processing, 22(6), 2296-2305.
[16] Bissacco, A., Cummins, M., Netzer, Y., & Neven, H. (2013). Photoocr: Reading text in
uncontrolled conditions. In Proceedings of the IEEE International Conference on Computer
Vision (pp. 785-792).
[17]Yin, X. C., Yin, X., Huang, K., & Hao, H. W. (2014). Robust text detection in natural
scene images. IEEE transactions on pattern analysis and machine intelligence, 36(5), 970-
983. using these references write the related work