IJCRT.ORG ISSN : 2320-2882

INTERNATIONAL JOURNAL OF CREATIVE RESEARCH THOUGHTS (IJCRT)

An International Open Access, Peer-reviewed, Refereed Journal

Face Recognition And Text Identification For Visually Impaired People

[1]Dr. I. Ambika, [2]Rudrakshi Tiwari, [3] S Sanathan Karthik, [4] S Sai Ashwin, [5] Sabah Alia R

[1]Associate Professor, [2]Undergraduate Scholar, [3]Undergraduate Scholar [4]Undergraduate Scholar, [5]Undergraduate Scholar

[1]Department of Computer Science and Engineering,

[1]MVJ College of Engineering, Bangalore, Karnataka, India.

Abstract: Today, visual information is the most important information, therefore blind or partially sighted people are at a disadvantage because mandatory avoidable information is not available to them. The aim of this project is to skillfully guide the visually impaired. This system is designed for them to solve an impossible situation that affects the blind. Our system introduces an automatic system that allows them to read text patterns printed on documents. Text models are localized and binarized using OCR. The identification text is converted to sound and given to the user. So work on character recognition. Difficulty recognizing people during a conversation is a huge handicap for those people in many professional and constructive situations. In these devices, smartphones are used to increase the accuracy of face recognition, we have an archetype and face recognition is tested on end users. This system uses a speech-to-speech module to generate audio for the blind.

Index Terms **– OpenCV, GTTS, Deep Learning, Pyttsx3, Tkinter, Text to speech**

I. INTRODUCTION

Low vision refers to a condition characterized by reduced visual acuity, ranging from partial to total blindness. Unfortunately, traditional solutions such as glasses or contact lenses cannot completely resolve the challenges posed by blindness. Losing one's sight also means losing a significant degree of independence. Reading plays a crucial role in our society, as it provides valuable information about various aspects of our lives and helps us navigate the world. Text can be found in diverse forms, including reports, receipts, product packages, and even restaurant menus.

For individuals with visual impairments, reading a text document poses a formidable challenge. Additionally, there are instances when a document must remain private, such as personal bank statements, tax invoices, or personal letters. However, due to the loss of vision, individuals have to rely on alternative aids to read these sensitive documents. In light of these considerations, there is a need for a system that can recognize text based on textual and symbolic data. This system would identify individual characters within an image of the text, subsequently converting the text or symbol data into a voice message. Several algorithms are employed to extract text information from images and facilitate text recognition.

Converting a landscape image into a textual format presents its own set of difficulties. Challenges include dealing with pixels at a specific level, distinguishing non-text background elements, and accommodating various foreground styles and sizes. Consequently, the process of extracting scene text involves two primary components: Optical Character Recognition (OCR) and Optical Character Retrieval (OCR). The former involves recognizing and extracting text from images, while the latter focuses on retrieving relevant textual information.

In addition to text identification, the system also includes an interface that can audibly communicate the names of buttons and other interactive elements. This feature greatly enhances accessibility and usability for visually impaired individuals. When interacting with digital devices, the system can detect the presence of buttons or interactive elements on the screen and provide voice prompts that inform the user about their names or functions. This empowers individuals with visual impairments to navigate user interfaces and interact with applications more effectively.

Furthermore, the system incorporates face recognition capabilities. By analyzing an image, the system should be able to recognize and match a face with the corresponding person's name in the training data. This feature enhances the system's utility in various applications, such as security systems or identification processes. By combining advanced algorithms and technologies, the aim is to develop comprehensive solutions that address the challenges faced by visually impaired individuals in both text identification and face recognition domains. These innovations can significantly improve accessibility, independence, and usability for individuals with low vision or blindness.

II. LITERATURE SURVEY

In this paper, they develop an object detection method that combines top-down detection and bottom-up image segmentation. This method has two main steps: a hypothesis generation step and a validation step. The top-down hypothesis generation step designs an improved shape context function that is more robust against object deformation and background clutter. The enhanced shape context is used to generate a set of hypotheses about object positions and shape ground masks with high recall and low accuracy rates. The validation step first computes a set of possible segmentations that match the top-down object hypotheses and then proposes a false positive pruning (FPP) procedure to exclude false positives. They take advantage of the fact that false-positive regions usually do not match the image segmentation potential. Experiments show that this simple framework can achieve both high recall and high accuracy on just a few positive training examples and that the method can be generalized to many object classes.[1].

Facial recognition is the ability to identify and recognize people based on their facial features. Faces are multidimensional and require a lot of mathematical calculations. Facial recognition systems are very important for providing security, facial photo matching, law enforcement applications, user authentication, user access control, etc., and are mainly used to recognize various applications. All these applications require efficient facial recognition systems. Many of the methods that have already been proposed have low detection capabilities and high false alarm rates. Therefore, the main research task is to develop a face recognition system with improved accuracy and recognition time. In this paper, he proposes a hybrid face recognition algorithm that combines two face recognition techniques through the integration of component analysis (PCA) and linear discriminant analysis (LDA). The Jacobi method is used to compute the eigenvectors required for the PCA and LDA algorithms. The facial recognition system is implemented on an embedded system-based Raspberry Pi 3 board.[2].

Visually impaired people face many challenges in their daily lives. in the last few years Innovative AI technology has been proposed and developed to support facial treatment for visually impaired people and object identification. These methods were mainly useful for visually impaired people to distinguish. People around you (using smart canes and glasses) understand and productively explore people's emotions. The outcome of this paper is to explore how these revolutionary advances can help. visually impaired people. Various techniques and algorithms. used to take necessary action software. PCA is one of many methods used to take a test image and contrast it against it. Images are stored in the information base. The OpenCV library was used to run the model. glass Models have also been created using Raspberry Pi, cameras, and sensors. of Appearance recognition is a multifactorial task, so techniques must be used to separate them We solve the problem by attacking various variables that directly affect the algorithm and its performance. [3].

A text is a system of signs used to record, transmit, and transmit culture. As one of mankind's most influential inventions, texts have played an important role in human life. In particular, rich and concise semantic information contained in the text is important in various vision-based application scenarios. With the advent and development of deep learning, computer vision has undergone great changes and transformations. As an important research area of computer vision, scene text recognition and recognition have become inevitable. Following this wave of revolution, we entered the era of deep learning. There have been significant advances in community thinking, methodology, and performance in recent years. The purpose of this study is to summarize and analyze the main changes and major developments in scene text detection and recognition in deep learning. [4].

This document identifies and compares different ones. Stages of text recognition and recognition process Analyze different approaches to text extraction color image. Two commonly used methods for this problem are a stepwise method and a built-in method, but this is Tasks are further divided into text recognition and localization. Classification, segmentation, text recognition. important Approaches are used to go through these phases and their stages. Corresponding advantages and disadvantages and possible uses Featured in this article. Various text-related applications images are also featured here. This review Performs a comparative analysis of basic processes in this field. [5].

Now text recognition is essential A problem for the visually impaired. text-to-speech (TTS) sentences. This article presents research and topics related to text recognition. Text-to-speech conversion. there was a lot of the text recognition method of the previous work. this paper Describes various text method techniques recognition [6].

Able to find and read text embedded in images of natural landscapes, the system is extremely useful for the visually impaired and partially sighted, providing useful information in everyday life and increasing self-confidence and autonomy. Optical character recognition (OCR) programs available today are fast and accurate, but most cannot recognize text embedded in images of natural scenes. The purpose of the algorithm described in this article is to identify text-like image regions and preprocess them in a way that allows OCR to work more reliably. The approach described in this work is based on the segmentation of color images and analysis of segment shapes. Preliminary tests show that the proposed algorithm provides a satisfactory detection rate and is fairly robust to typical text distortions such as tilt, tilt, and curvature. [7].

Text recognition in natural and complex images plays an important role in image analysis. Textual information appears everywhere, including product labels, documents, and manufacturing photos. Visually impaired people find it very difficult to find text areas in images. We cannot assume that the captured image contains only text, so we have to deal with this issue. This article introduces a camera-based text-reading framework that helps visually impaired people read text such as natural landscapes and product labels. The task here is divided into two steps: text recognition and text recognition. Text recognition includes a text localization phase. First, the image is binarized. Geometry and stroke weight filtering remove non-text areas from the image. Segmented text regions in crowded scenes are binarized and recognized by optical character recognition. Recognized text is spoken out for the visually impaired. Experimental results show that the proposed method provides better performance in text recognition.[8].

This article introduces intelligent bots for the assistance visually impaired. Currently, 81% are visually impaired people living in developing countries. modern human Communication will focus primarily on text and voice. to read Text: People need vision. Conduct a survey on multiple papers and the system provides the hardware consisting of the camera interface Process text using Raspberry Pi. The camera captures A text image of handwritten or printed text. raspberry pi Uses optical character recognition (OCR) software. is installed to convert images to text, The same goes for text-to-speech conversion. The wizard can apply Suitable for both visually impaired and non-disabled people Increase comfort. [9].

III. WORKING PRINCIPLE

The Prototype is designed to extract text and recognize face and convert it to speech. Face recognition systems are mainly used to authenticate a person's information and Text identification systems are utilized to extract text from the document. Face recognition system has proven invaluable in matters of national security and law enforcement. Government agencies utilize face recognition systems to identify and track individuals of interest, aiding in the identification of suspects, locating missing persons, and ensuring public safety integrity and authenticity. On the other hand, Text recognition systems are being used by the government as well as private industries for document verification, ensuring the authenticity and integrity of important records such as passports, driver's licenses, and identity documents. The process of Face recognition and Text Identification involves a series of steps. Firstly, in Face recognition, there should be enough data collected for a person by collecting at least 300 images. For text identification text is acquired digitally then the text from the documents is extracted. After the image and the text are extracted, they undergo pre-processing to enhance their quality and eliminate any distortions. Techniques like noise removal and image enhancement are applied during this stage.

Following pre-processing, features are extracted from the images and the documents. For face recognition characteristics such as facial landmark detection, local feature extraction, texture analysis, and statistical analysis providing information about the general shape and structure of the face are extracted. Simultaneously, a reference database is created, containing a genuine face for comparison. This database is populated during the enrollment process, where individuals provide their genuine faces for future verification purposes. On the other hand, text identification focus on specific details within the document, such as shape, width texture, gradient histograms, and statistical properties. The extracted features are then utilized to represent the text in a numerical or mathematical form, often in the form of feature vectors or templates.

For face recognition, the face to be verified is compared with the reference face within the database using various comparison techniques, such as distance-based metrics or deep learning algorithms. This technique analyzes inconsistencies between the questioned face and genuine faces, considering factors like facial landmark detection, local feature extraction, texture analysis, and statistical analysis. On the other hand, text identification techniques are used to extract the text from the document. Based on the results of shape, width texture, gradient histograms, and statistical properties. Then the extracted text would be converted into speech by using the correct modules.

Face recognition systems can leverage Convolutional Neural Networks (CNNs) to analyze and classify facial images and text documents. CNNs, designed for processing grid-like data like images, can learn features from identified and unidentified face datasets and text. The trained CNN model extracts features from identified and genuine face datasets as a reference database. Additionally, CNNs can detect forgery by analyzing inconsistencies and irregularities in facial features. CNNs enhance system efficiency by automatically learning and extracting relevant features, improving the overall effectiveness of facial recognition.

Optical Character Recognition (OCR) technology is employed in text identification to extract textual information from images or documents. The OCR process involves several steps. First, the input image or document is acquired through scanning or digital capture. Pre-processing techniques are applied to enhance the image quality, such as noise removal, binarization, and deskewing. Next, the image is segmented into individual characters or text regions. This segmentation allows for the isolation and identification of each textual element. Feature extraction techniques are then utilized to capture important characteristics of the characters, such as shape, stroke width, texture, and statistical properties. These features are encoded into a numerical representation.

Classification algorithms or machine learning models are applied to recognize and interpret the characters, mapping them to their corresponding textual values. Finally, the recognized text is outputted, providing the user with the extracted information. OCR technology finds applications in various domains, including document digitization, data entry automation, and accessibility for visually impaired individuals, playing a vital role in enabling efficient text identification and enhancing information retrieval processes.

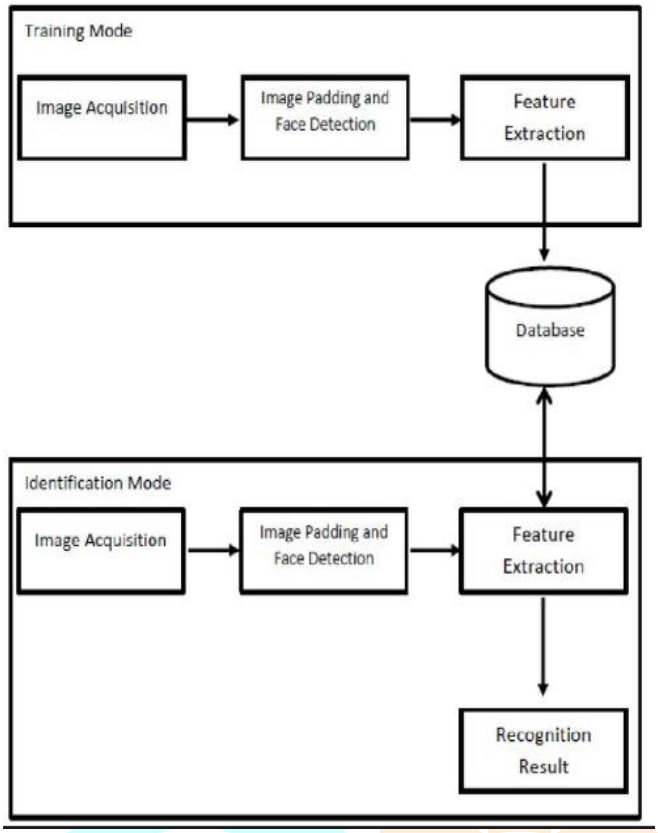

Figure 1: Workflow of the Face Recognition

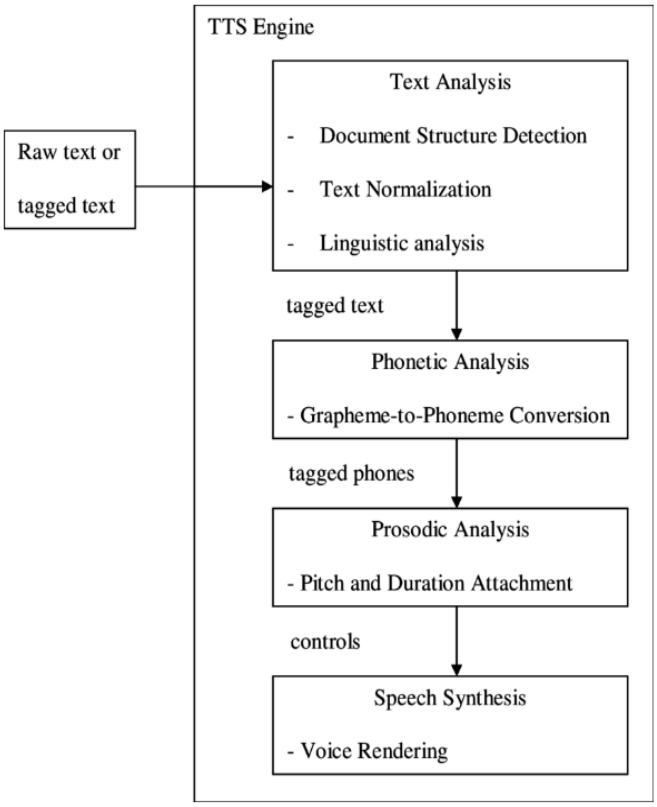

Figure 2: Workflow of the Text Identification

IV. IMPLEMENTATION

This prototype utilizes different concepts at different stages to achieve the needful. Firstly, we have used a simple menu divine program which lets the user choose if they want to use face recognition or text identification. If the user chooses face recognition, then the deep learning model first created a dataset for the user to register its name. In the dataset, at least 300 images of the user are stored and the model is trained on these 300 images. Then after creating the dataset the training starts, and the model learns to minimize a defined loss function by adjusting its weights using optimization algorithms. Once trained, the model can extract discriminative features from face images. To recognize a face, the extracted features are compared to a database of known faces using similarity metrics, with a threshold determining the match. On the other hand, if the user chooses the Text identification OCR algorithm to extract the text from the document. OCR algorithms are applied to extract textual information from documents or images that contain signatures. This is done by scanning physical documents, capturing digital images, or targeting specific regions of interest within an image. The OCR feature allows the system to extract text from the image, enabling the detection of specific keywords or information such as IFSC codes or predefined keywords like "Please" or "Sign." This extraction process is achieved using the Tesseract OCR engine, which recognizes text in documents. The OCR system analyzes the visual characteristics of the text, including the shapes of individual characters and their spatial relationships, to convert them into machine-readable text. Then the extracted text gets converted into speech.

OCR (Optical Character Recognition) in text identification works by using computer vision techniques and machine learning algorithms to extract text information from images or scanned documents. The process involves several steps. First, the input image is preprocessed to enhance its quality, correct distortions, and normalize the text. Then, the image is segmented into individual characters or text regions. Next, feature extraction techniques are applied to represent the segmented text regions in a suitable format for recognition. These features can include the shape, texture, and statistical properties of the characters. Machine learning algorithms, such as neural networks or support vector machines, are trained on a dataset of labeled text samples to learn the patterns and variations in different characters or fonts.

During recognition, the extracted features are compared to the learned models to determine the most likely characters or words. This process involves classification algorithms that assign probabilities or confidence scores to each candidate character or word. The highest-scoring candidates are selected as the recognized text.OCR systems may also incorporate language models or dictionaries to improve recognition accuracy by considering contextual information and word patterns. Additionally, post-processing techniques like error correction and spell-checking can be applied to refine the recognized text. Overall, OCR in text identification enables the automated extraction of text from images or scanned documents, making it valuable in various applications such as document digitization, text extraction from images, and accessibility for visually impaired individuals.

Text identification and face recognition are two important areas of computer vision and pattern recognition. In text identification, the process typically involves acquiring an image or document containing text, preprocessing the image to enhance its quality, and then extracting the text using optical character recognition (OCR) techniques. OCR algorithms analyze the image, identify individual characters, and convert them into machine-readable text. This extracted text can then be further processed for tasks such as text recognition, language modeling, and information retrieval.

On the other hand, face recognition focuses on detecting and identifying human faces in images or video frames. The process begins with face detection, where algorithms locate and localize facial regions within the input data. Once the faces are detected, features are extracted from these regions to represent each face uniquely. Traditional methods use techniques like eigenfaces, LBPs, or HOG to extract features, while deep learning approaches employ convolutional neural networks (CNNs) to learn discriminative features directly from raw image pixels. These extracted features are then encoded into compact representations called face embeddings or feature vectors.

During the recognition phase, the system matches the face embeddings of the input face with known face embeddings stored in a database. This comparison is done by measuring the similarity or distance between the embeddings using metrics like Euclidean distance or cosine similarity. If the distance falls below a certain threshold, it signifies a match, and the input face is recognized as belonging to a known individual.

Both text identification and face recognition rely on advanced algorithms and techniques, including image processing, machine learning, and pattern recognition. The development of deep learning models has significantly improved the accuracy and performance of these systems. These technologies find diverse applications in fields such as security, biometrics, document processing, and accessibility for visually impaired individuals. Continued advancements in these areas hold the potential for further enhancing the accuracy and robustness of text identification and face recognition systems.

When a face is detected and recognized, the system can convert the identified name into speech using TTS, allowing visually impaired users to hear and perceive the name associated with the recognized face.

In this process, once the face recognition algorithm successfully identifies a person, the corresponding name is obtained from the database or recognition model. The name is then passed to the TTS component, which converts the text representation of the name into audible speech. This synthesized speech can be played through speakers, headphones, or any audio output device to convey the name to the visually impaired user. The TTS system employs various techniques, such as linguistic analysis, language modeling, and voice generation, to ensure that the synthesized speech is clear, natural-sounding, and easily understandable. The system may consider factors like pronunciation rules, intonation, and voice characteristics to enhance the overall listening experience.

In text identification when the system successfully identifies and extracts text from an image or document, TTS is employed to convert the text into spoken words, allowing visually impaired users to access and comprehend the content. In the text identification process, optical character recognition (OCR) techniques are utilized to recognize and extract text from various sources such as

scanned documents, images, or screenshots. Once the text is extracted, it is passed to the TTS component, which converts the textual information into synthesized speech.

The TTS system utilizes natural language processing techniques and speech synthesis algorithms to generate human-like speech. It analyzes the text, considers factors like grammar, punctuation, and sentence structure, and applies intonation and prosody rules to produce speech that closely resembles natural human speech patterns.

The synthesized speech output from the TTS system can be played through speakers, headphones, or any audio output device, allowing visually impaired individuals to listen to the content of the identified text. This enables them to access printed materials, digital documents, or any textual information that would otherwise be inaccessible without visual assistance.

By integrating TTS into text identification systems, visually impaired individuals gain the ability to interact with and consume textual content independently. It empowers them to access a wide range of information, including books, articles, product labels, and more, enhancing their overall autonomy and inclusivity in the digital and print world.

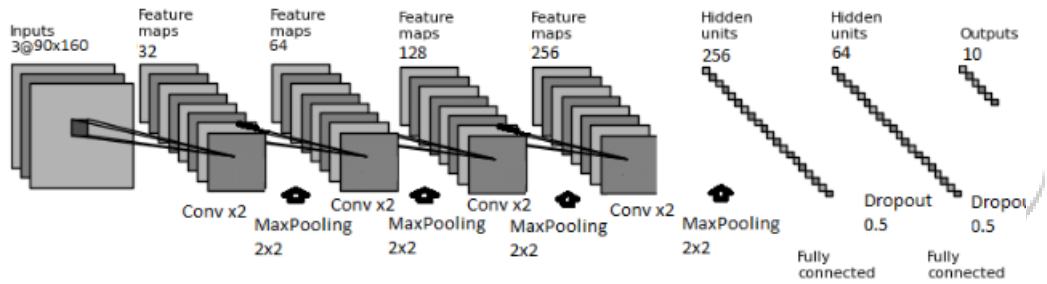

Figure 3: CNN Model for face recognition

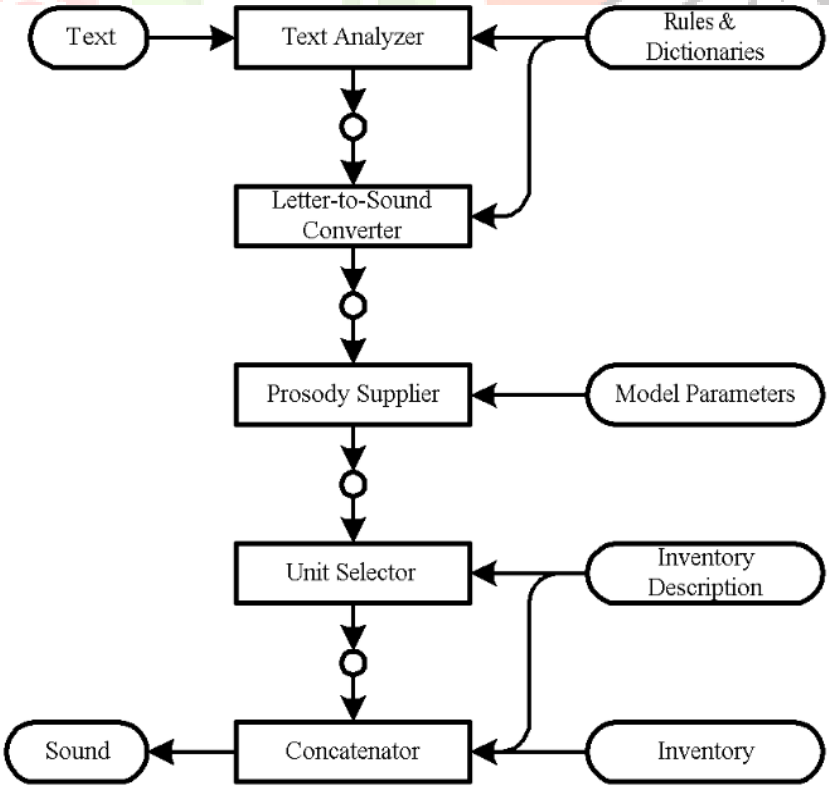

Figure 4: Working for Text Identification

V. RESULTS

This is the graphical user interface (GUI) for Face Recognition and Text Identification, it is a simple and easy-to-use application; where not much training is required. The GUI is intuitive and user-friendly, allowing users to navigate and interact with the system effortlessly. The GUI has a clean and organized layout, featuring clear instructions and guidance at each step of face recognition and text identification. Users can easily upload samples, and interpret the results. The system will leverage user-centered design principles, conducting usability tests and incorporating user feedback to optimize the interface for an efficient and enjoyable user experience.

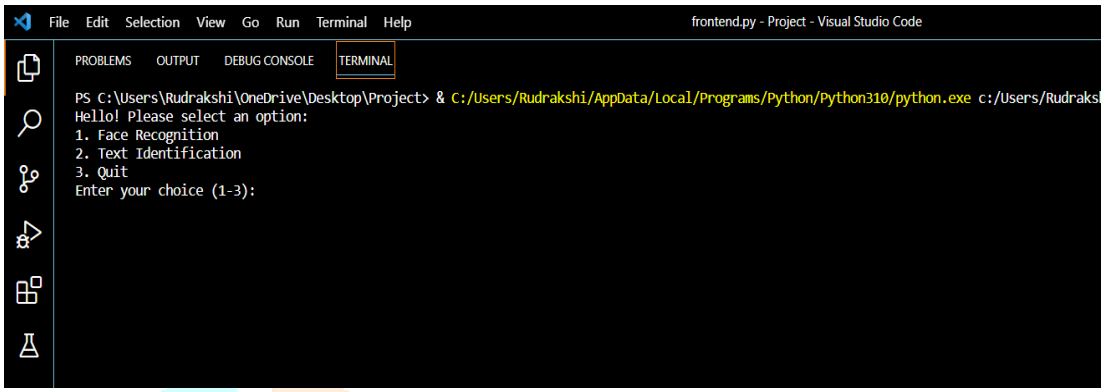

Figure 5: Starting menu

After choosing any one option the program will be directed to the specific page. If the user chooses Face recognition then the program opens this prompt.

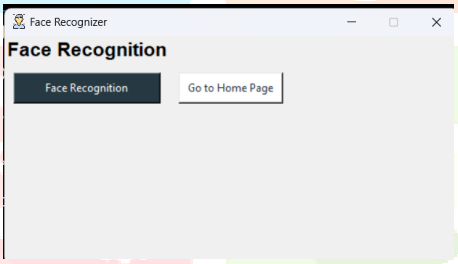

Figure 6: Face Recognition Starting Page

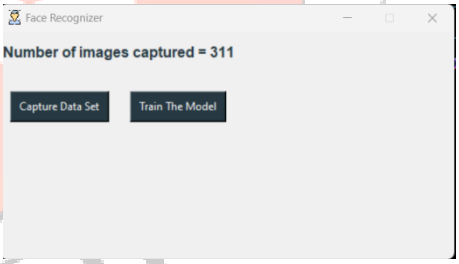

Figure 7: Face Recognition Making Dataset

After clicking on Face recognition, the system will first check if the user is registered or not, if not registered then it will capture at least 300 pictures. After capturing we can train the model and the system will verify the user.

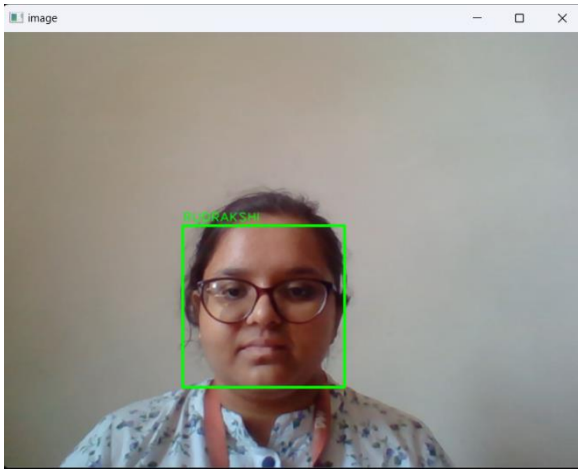

Figure 8: Detecting Faces

To leave the program the user can press and hold ‘q’ for a while and the program will quit the face recognition window. Then the user can choose Text identification from the menu option. The text recognition window will open, the user would then have to select a pdf to be converted to speech and press the Convert to Speech button in the window.

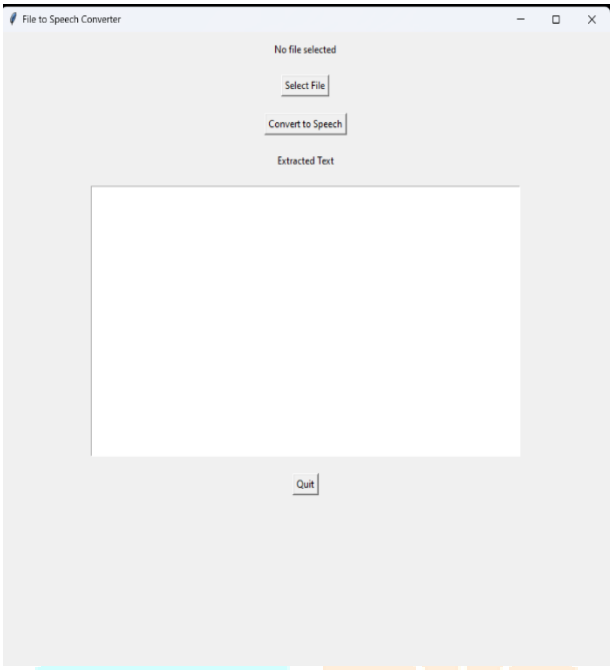

Figure 9: Text Identification Page

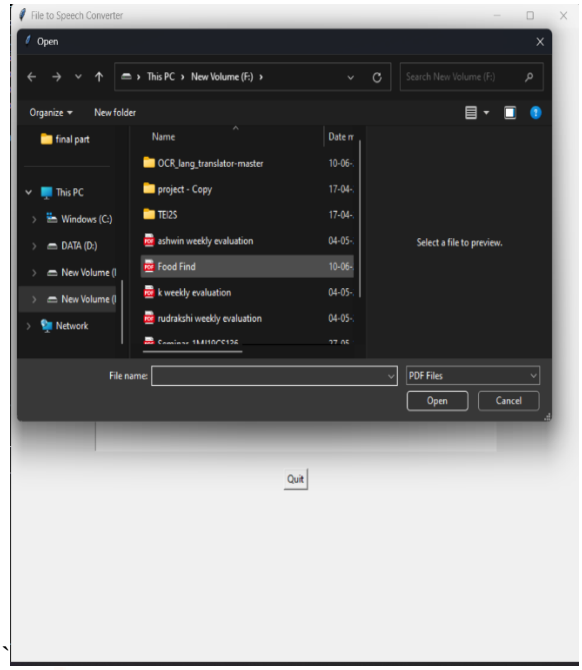

Figure 10: Selecting a file

After pressing the button, the text will be converted to speech and the text extracted will open in the text box

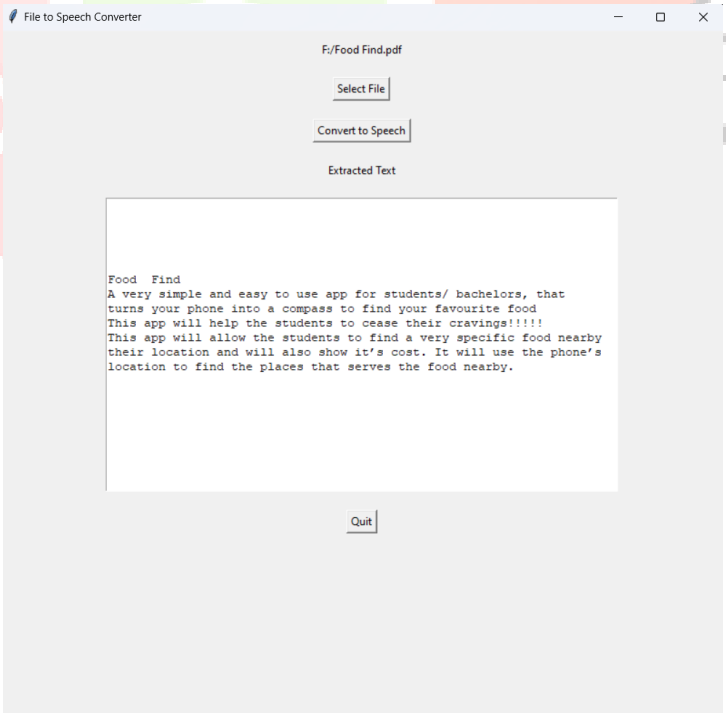

Figure 11: Text converted to speech

The user can quit the program by pressing the quit button in the window. To quit the whole program the user would have to choose the quit option in the python prompt.

VI. FUTURE SCOPE

[1] One area of focus will be enhancing the accuracy of face recognition and text identification systems. This can involve advancements in deep learning models, feature extraction techniques, and algorithm optimization to achieve more reliable and precise results.

[2] Real-time and Robust Performance: There is a growing demand for real-time face recognition and text identification systems that can operate in dynamic and challenging environments. Future developments may involve optimizing algorithms and hardware to achieve faster processing speeds while maintaining high accuracy and robustness against variations in lighting, pose, and occlusions.

[3] Future advancements in face recognition and text identification may involve exploring novel neural network architectures, leveraging generative models, or incorporating advanced techniques like self-supervised learning to improve performance.

[4] Integrating Internet of Things (IoT) devices and edge computing capabilities will allow distributed processing, reduced latency, and improved privacy, enabling applications in smart surveillance, access control, and personalized services.

VII. CONCLUSION

As technology keeps on progressing at a very fast pace, making sure that this technology is affordable and reachable to all types of people is important. So as the new technology keeps on developing for the visually impaired and making sure the gadgets that are being used are up to date. Text recognition and Face identification is developed to deal with people with visual disability, with the help of this system the person can hear text that is being identified in the document and can also be able to hear the name of the person that is recognized in the webcam. Over going through many research papers, we have found many studies and models made using OpenCV and CNN neural networks to be more optimum in their working and more accurate. Going forward we are using these two as the main algorithm and libraries to build the foundation for the project.

REFERENCES

[1] Object Recognition**:** Liming Wang, Jianboa Shi, Gang Song, and I-fan Sheen

[2] Real-time Implementation of Face Recognition System- Pradeepa .M, H P Mohan Kumar

[3] REVIEW PAPER ON FACIAL RECOGNITION FOR VISUALLY IMPAIRED PEOPLE Dev Gupta, Shourya Gupta, Aditi Khurana

[4] A Study of Text Detection and Recognition: Neel Ramakant Borkar; Sonia Kuwelkar

[5] Text Detection and Recognition: A Review -Chaitanya R. Kulkarni, Ashwini B. Barbadekar

[6] A REVIEW ON TEXT RECOGNITION FOR VISUALLY BLIND PEOPLE-Siva Ganga , Hema S

[7] Text Detection System for the blind: Marcin Pazio, Maciej Nied´zwiecki, Ryszard Kowalik, Jacek Lebied´z

[8] Text Recognition System for Visually Impaired**:** Gauri Vaidya, Ketki Vaidhya, Kishor Bhosale

[9] A Smart Reader for Blind People -L Latha, V Geethani, M Divyadharshini, P Thanga