AUDIOBOOK FOR VISUALLY IMPAIRED PEOPLE

[1]Prof. Prakruthi G R, [2]Jashwanth P V, [3]A M Pavan, [4]Deekshith K S, [5]Neeraj Y S

Department of Information Science and Engineering

East West Institute of Technology, Bangalore, India

Abstract: There are close to 39 million blind people and around 285 million visually impaired people globally. There is a huge impact on the lives of visually disabled people due to this. Although there have been several attempts made for helping visually disabled to see objects via other alternating means such as sound and touch, the development of text reading device is still at a nascent stage. The system currently in existence either has a limited scope or requires a heavy investment. Therefore, we need a cost effective and truly efficient system that will be able to automatically identify and recite text aloud to visually challenged user base.

Index Terms - **Audiobook, tessract ocr, espeak sythesizer, TTS(Text-to-speech), PDF, image.**

I. INTRODUCTION

Machine learning techniques one of the most precious gifts to a human being is an ability to see, listen, speak and respond according to the situations. But there are some unfortunate ones who are deprived of this. Making a single compact device for people with Visual impairment is a tough job. We provide a technique for a blind person to read a text and it can be achieved by capturing an image through a camera that converts a text to speech (TTS). The blind people can be able to read the words using by Tesseract OCR (Online Character Recognition) read out by espeak, all these functions are implemented by the use of Laptop.

Image to speech conversion is a trending aspect of computer technology. It determines an important criterion in which we interact with the system and interfaces across a variety of platforms. ¬ Machine replication of human functions, like reading, is an ancient dream. However, over the last five decades, machine reading has grown from a dream to reality. ¬ Speech is probably the most efficient medium for communication between humans. Optical character recognition has become one of the most successful applications of technology in the field of pattern recognition and artificial intelligence. An audiobook is a recorded version of a book, which can be played on a variety of devices such as smartphones, tablets, laptops, and dedicated audiobook players.

Audiobooks have become increasingly popular in recent years as they provide an accessible way for visually impaired individuals to enjoy books, whether for entertainment or education. Many publishers and organizations now offer audiobook versions of their books, which are often narrated by professional actors or voice artists to bring the text to life.

Audiobooks offer a range of benefits for visually impaired people. They allow readers to access books that may not be available in braille or large print formats, and they provide a convenient way to consume books without the need for physical storage space. Audiobooks can also be listened to on the go, making it possible to enjoy books during activities such as exercise, commuting, or doing chores. It is an optical character recognition engine for various operating systems. Tesseract up to and including version 2 could only accept TIFF images of simple one-column text as inputs. These early versions did not include layout analysis, and so inputting multi-columned text, images, or equations produced garbled output. Since version 3.00 Tesseract has supported output text formatting, OCR positional information and page-layout analysis. Support for a number of new image formats was added using the Leptonica library. Tesseract can detect whether text is mono spaced or proportionally spaced.

Text recognition and extraction is needed when the information should be readable both to humans and to a machine and alternative inputs cannot be predefined. The basic Text extraction system was invented to convert the data available on papers in to computer process able documents, so that the documents can be editable and reusable. Traditional techniques are typically multi-stage processes. For example, first the image may be divided into smaller regions that contain the individual characters, second the individual characters are recognized, and finally the result is pieced back together. A difficulty with this approach is to obtain a good division of the original image. Though tremendous strides have been made in character recognition but it is still considered to be a difficult problem when the data is rotated and non-uniform in scale. We have seen that very few works have been done for

Indian languages using OCR. In this work, we have taken the problem of improving the recognizing capability of compound characters using OCR so as to achieve accurate character values.

II. RELATED WORKS

A chat-bot is a software tool that uses text or text-to-speech to conduct an online chat conversation instead of providing direct contact with a live human agent. The proposed system in this paper learned the Chatbot by watching numerous user responses and requests and utilising a keyword matching technique using machine learning techniques [1]. This prototype chatbot analyses the user context to prompt a response with a specified intent. Because it is a dynamic response, the user will receive the desired response. This type of web-based platform provides a large cognitive base that can be used to replicate human problem-solving.

The introduction of AI into the banking industry has spurred chatbots and changed the face of bank-customer contact. The most recent disruptive force that has revolutionised how customers connect is the rise of chatbots in the banking and finance industries. This study presents a new approach for dealing with artificial intelligence in addition to highlighting the capabilities of intelligent systems. The banking sector is crucial to any country's development [2]. It also looks into the chatbot's current usability to see if it can meet the ever-changing needs of customers.

In one of the previous study, the authors will discuss the role of chatbots in education and e-commerce, as well as describe chatbot classification and techniques. Chatbots are becoming increasingly popular in business, necessitating the implementation of novel approaches to providing 24-hour customer service. This type of business is especially important in these difficult Covid-19 times and its after affects. Chatbots powered by artificial intelligence can function as intelligent teaching systems, providing users with a personalized learning experience [3].

III. PROPOSED SYSTEM

Our proposed system is Extraction of text from an image using OCR, on a grid infrastructure which is a character recognition system that supports recognition of the characters of multiple languages. This feature is what we call grid infrastructure which eliminates the problem of heterogeneous character recognition and supports multiple functionalities to be performed on the image. In this context, Grid infrastructure means the infrastructure that supports group of specific set of languages. Thus, Extraction of text from an image using OCR on a grid infrastructure is multi-lingual.

The benefit of proposed system that overcomes the drawback of the existing system is that it supports multiple functionalities such as editing and searching. It also adds benefit by providing heterogeneous characters recognition. This system recognizes the characters approximately based on their trained data values. OCR is used to match the recognized character with the trained dataset storage.

IV. IMPLEMENTATION

The proposed method is applied on gray scale image as well as color image. Here the curled text lines are detected by using the properties of connected components. Then the text is extracted from the document image using OCR. It can be used further for document image analysis, understanding and processing. Equation describes a line with slope and β intercept y, α.Segmentation is the most important aspect in pre-processing. From each individual character it extracts the features for character recognition. Moments that make the process of recognizing an object scale, translation, and rotation invariant can be used for character recognition. Text detection refers to the determination of the presence of text in a given sequence of images. Text localization is the process of determining the location of text in the image and generating bounding boxes around the text. Text tracking is performed to reduce the processing time for text localization and to maintain the integrity of position across adjacent frames. Although the precise location of text in an image can be indicated by bounding boxes, the text needs to be segmented from the background to facilitate its recognition. That means, the extracted text image has to be converted into a binary image and enhanced before it is fed into an OCR engine. Text extraction is the stage where the text components are segmented from the background.

Text Extraction from image is concerned with extracting the relevant text data from a collection of images. Recent studies in the field of image processing show a great amount of interest in content retrieval from images. This content can be in the form of objects, color, texture, shape as well as the relationships between them. The semantic information provided by an image can be useful for content based image retrieval, as well as for indexing and classification purposes. Since the text data can be embedded in an image in different font styles, sizes, orientations, colors, and against a complex background, the problem of extracting the candidate text region becomes a challenging one. Current Optical Character Recognition (OCR) techniques can only handle text against a plain monochrome background and cannot extract text from a complex or textured background. AText Information Extraction system receives an input in the form of a still image or a sequence of images. The images can be in gray scale or color, compressed or uncompressed, and the text in the images may move or may not.
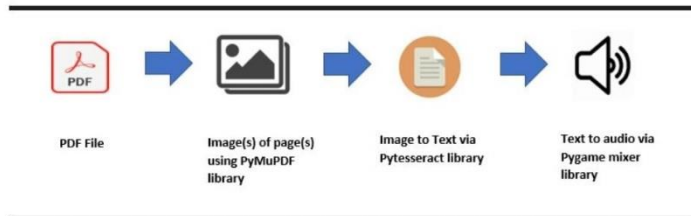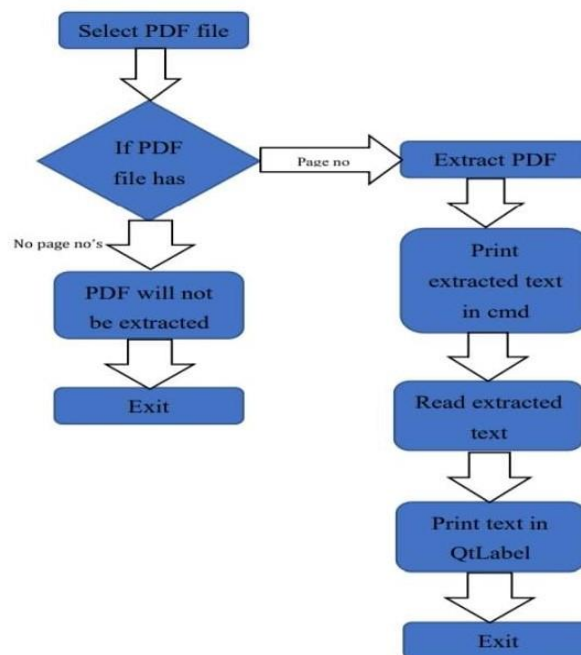
Figure 1: Audiobook architectutre

Figure 2: Flow Chart

OCR (Optical Character Recognition)

Optical character recognition or optical character reader (OCR) is the mechanical or electronic conversion of images of typed, handwritten or printed text into machine- encoded text.
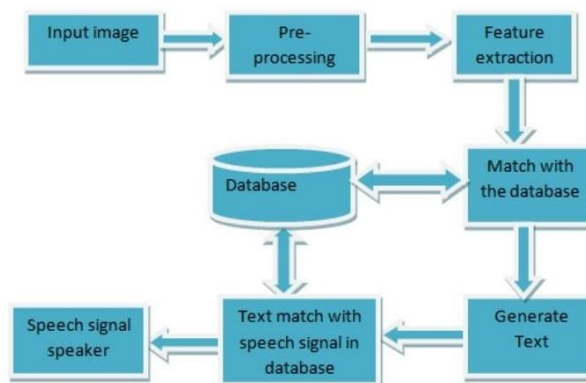

Figure 3: OCR Architecture

Pre-processing

a) De-skew: To make lines of text perfectly horizontal or vertical.

b) De-speckle: Remove positive and negative spots, smoothing edges.

c) Binarisation: Convert an image from color or gray scale to black-and-white.

d) Line removal: Cleans up non-glyph boxes and lines

e) Layout analysis: Identifies columns, paragraphs, captions, etc. as distinct blocks.

f) Line and word detection: Establishes baseline for word and character shapes, separates words if necessary.

g) Script recognition: The script may change at the level of the words and hence, identification of the script is necessary.

h) Character isolation: For per-character OCR, multiple characters that are connected due to image artifacts must be separated; single characters that are broken into multiple pieces due to artifacts must be connected.

An algorithm is used for organizing the extracted text. The steps involved in the algorithm is as follows:

Step1: All the Extracted text is first analysed, then each and every text has a particular format i.e pattern.

Step2: Identify the pattern for each and every line of the text occurred. Consider the Name, for identifying the text as in the name all the characters should be alphabets but not numbers. Similarly, For identifying the text as phone number it should contain 10 digits, then it is named as phone number or contact number.

Step3: All the useful information is organized in the similar pattern recognition format. The rest is unaltered/ not considered.

Step4: Output of the organized text is printed.

Step5: The text will be given in voice output using text to speech libraries.

V. RESULTS

A digital library of audio books that can be accessed online or through a mobile app. The library should have a wide range of books covering various genres and categories. areas which are capable of operating without human intervention at any time regardless of the location.

Figure 1:Audiobook choices

A user-friendly interface that visually impaired individuals can easily navigate using assistive technology. This interface should allow users to search for and select the books they want to listen to.Text-to-Speech Conversion is a software system that can convert text from a book into spoken words. This system should be able to read aloud the entire text of a book in a clear and natural-sounding voice. Book Navigation is an intuitive navigation system that allows users to easily move between chapters, sections, and pages of the book. This navigation system should be accessible through voice commands and assistive technology prediction.

The proposed method is applied on gray scale image as well as color image. Here the curled text lines are detected by using the properties of connected components. Then the text is extracted from the document image using OCR. It can be used further for document image analysis, understanding and processing. Equation describes a line with slope and β intercept y, α.Segmentation is the most important aspect in pre-processing. From each individual character it extracts the features for character recognition. Moments that make the process of recognizing an object scale, translation, and rotation invariant can be used for character recognition. Text detection refers to the determination of the presence of text in a given sequence of images. Text localization is the process of determining the location of text in the image and generating bounding boxes around the text. Text tracking is performedto reduce the processing time for text localization and to maintain the integrity of position across adjacent frames. Although the precise location of text in an image can be indicated by bounding boxes, the text needs to be segmented from the background tofacilitate its recognition. That means, the extracted text image has to be converted into abinary image and enhanced before it is fed into an OCR engine. Text extraction is the stage where the text components are segmented from the background.

Text Extraction from image is concerned with extracting the relevant text data from a collection of images. Recent studies in the field of image processing show a great amount of interest in content retrieval from images. This content can be in the form of objects, color, texture, shape as well as the relationships between them. The semantic information provided by an image can be useful for content based image retrieval, as well as for indexing and classification purposes. Since the text data can be embedded in an image in different font styles, sizes, orientations, colors, and against a complex background, the problem of extracting the candidate text region becomes a challenging one. Current Optical Character Recognition (OCR) techniques can only handle text against a plain monochrome background and cannot extract text from a complex or textured background. A Text Information Extraction system receives an input in the form of a still image or a sequence of images. The images can be in grayscale or color, compressed or uncompressed, and the text in the images may move or may not.

REFERENCES

[1] L. Neumann and J. Matas. A method for text localization and recognition in real-world images. In ACCV, pages 770–783, 2010.

[2] Reza Sarshogh and Keegan Hines, "Computer Vision Methods for Extracting Text from Images", Capital One Tech

[3] T.Som, Sumit Saha,"Handwritten Character Recognition Using Fuzzy Membership Function", International Journal of Emerging Technologies in Sciences and Engineering, Volume 5, December 2011.

[4] L. A. Zadeh. Fuzzy sets, Information Control 8 (1965) 338–353.

[5] Prakhar Gur, Eran, and ZeevZelavsky, "Retrieval of Rashi Semi-Cursive Handwriting via Fuzzy Logic," IEEE International Conference on Frontiers in Handwriting Recognition (ICFHR), 2012

[6] Thomas Natsvhlager, "Optical Character Recognition", A Tutorial for the Course Computational Intelligence.

[7] Andrei Polzounov, ArtsiomAblavatski , Sergio Escalera, Shijian Lu, JianfeiCai "Wordfence: Text Detection In Natural Images With Border Awareness" ,2020.

[8] Johanes effendi, "End-to-End Image-to-Speech Generation for Untranscribed Unknown Languages", 2021

[9] Rithika.H , B. Nithya santhoshi, "Image Text To Speech Conversion In The Desired Language By Translating With Raspberry Pi", 2020.

[10] Sneha.C.Madre, Prof.S.B.Gundre, "OCR Based Image Text To Speech Conversion Using MATLAB", 2019.

[11] Nikolaos Bourbakis, "Automatic Image-to-Text-to-Voice Conversion for Interactively Locating Objects in Home Environments", 2018.

[12] Garibay, Cesar Millan Ollivarria, "MyVox—Device for the communication between people: blind, deaf, deaf- blind and unimpaired", 2017.

[13] SAKRIANI SAKTI , (Member, IEEE), AND SATOSHI NAKAMURA, "End-to-End Image To Speech Conversion", 2016.