International Journal of Multidisciplinary Research and Growth Evaluation
ISSN: 2582-7138
Received: 20-09-2021; Accepted: 26-10-2021
www.allmultidisciplinaryjournal.com
Volume 2; Issue 6; November-December 2021; Page No. 364-373

Text Reader

Dr. M Upendra Kumar [1*], Imtiyaz khan [2], Juveriya Nishath [3], Afrah Rakshan [4], Fabiha Fathima Unsa [5]
[1] Professor of CSE MJCET OU Hyderabad, Telangana, India
[2] Assistant Professor CSE MJCET OU Hyderabad, Telangana, India
[3-5] BE CSE MJCET OU Hyderabad, Telangana, India

Corresponding Author: **Dr. M Upendra Kumar**

DOI: https://doi.org/10.54660/.IJMRGE.2021.2.6.364-373

Abstract
The historical and theoretical bases of contemporary high-performance text-to-speech (TTS) systems and their current designs are discussed.
The major elements of TTS system are described, with particular reference to vocal tract models.
The stages involved in the process of converting text into speech parameters are examined, covering text normalization, word pronunciation, prosodies, phonetic rules, voice tables, and hardware implementation.
It is an image to text and speech conversion system developed for visually impaired as well as physically challenging people to be able to get information from the images easily.
Core idea for image to text and speech conversion is to overcome the challenges faced by a visually impaired person in real life. The system goes through various phases such as image processing, text extraction, text-to-speech (TTS) conversion.
This device opens the camera using the app into it out to the text and you get the audio.
The primary motivation is to provide a visually impaired person with a friendly speech interface with computer and to allow such people who are physically and visually challenged to use the system to read printed text on the go.

Keywords: Text, Reader, TTS

Introduction
Reading is very important and one aspect in our day to day life. Almost 314 million visually impaired people are there all around the universe, 45 million are blind and new cases being added each year as per the researches done. For example, by 2050 there could be 115 million people who are blind, up from 38.5 million in 2020.
Emerging technologies and recent developments in computerized vision, digital cameras, PDA and portable computers make it flexible to assist these individuals by developing image-based products that combine computer vision technology with other existing commercial technology such as optical character recognition. (OCR) platforms.
Printed text is one of the forms of reports, receipts, bank statements, restaurant menus, classroom handouts, product packages, medicine bottles, banners on road etc. For example, portable bar code readers designed to favor the blind people recognize dissimilar products, it permits the users who are unsighted to ingress information about these products through speech and Braille. But a big limitation is that it is very tough for unsighted people to find the location of the bar code and to exact point the bar code reader at the bar code.
There are systems like K Reader Mobile. It runs on a cell phone and permits the user to check mail, receipts, and many other things.
But, it only (K Reader Mobile) sees black print on a white background.
"Extraction of text regions in Natural images," the algorithm is robust in most common cases, but doesn't detect text, when it is in small size.
Accessing text documents is troublesome for visually impaired people in may scenarios, such as reading text on the go and accessing text in less than ideal conditions. The available technologies, such as screen readers, desktop scanners, smart phone applications, -book readers and embossers, are commonly under-utilized due to slow processing speeds or poor accuracy.
Technological barriers inhibit Visually impaired people's abilities to gain more independence, a characteristic widely identified as important.

The user scans the text and image is captured then text extraction is performed on it. For the extracted text, corresponding audio is generated which is available to the user.

Problem statement

Without vision it can be challenging for visually impaired person to read text from any printed material i.e, newspaper, books etc. Even with existing technologies, VI people face many difficulties including problems with alignment, focus, accuracy, mobility and efficiency in accessing printed text.

Objective

The aim of the project is to overcome the difficulties faced by the visually impaired and provide a system with speech assistance which is used to read any printed text from the newspapers, books etc. Our system scans the printed text by capturing image and extracts text from the image later converting the extracted text into audio.

Literature survey

In March 2015, the Massachusetts Institute of Technology (MIT) announced the the researchers in its Media Lab had developed a prototype of a reding device that is worn on the finger.

Many people in the accessibility community were very excited by this prospect. Unlike other common OCR (optical character recognition) apps that first scan and then process the page, the MIT device, dubbed the Finger Reader, reads the text in real time. The concept for the device was developed by Roy Shilkrot, an MIT graduate student in Media Arts and Sciences. He and Media Lab postdoc Jochen Maes, the Alexander W. Dreyfoos Professor in Media Arts and Sciences at MIT.

Suranga Nanayakkara, an assistant professor of engineering product development at the Singapore University of Technology And Design, who was a postdoc and later a visiting professor in Mae's lab, and MengEe Wong of Nanyang Technological University in Singapore Dr. Huber presented the paper in April at the Association for Computing Machinery Computr-Human Interface coference.

Researchers at the MIT Media Laboratory have built a prototype of a finger-mounted device with a built-in camera that converts written text into audio for visually impaired users. The device provides feedback - either tactile or audible - that guides the user's finger along a line of text, and the system generates the corresponding audio in real time.

The FingerReader is a wearable device that assists in reading printed text. It is a tool both for visually impaired people that require help with accessing printed text, as well as an aid for language translation.

Wearers scan a text line with their finger and receive an audio feedback of the words and a haptic feedback of the layout: start and end of line, new line, and other cues. The FingerReader algorithm knows to detect and give feedback when the user veers away from the baseline of the text, and helps them maintain a straight scanning motion within the line.

[1], Xiaodong Yang, VingliTianChucai Vi Aries Ardith "Context-based Indoor Object Detection as an aid to blind persons accessing Unfamiliar Environments" 2010 computer vision based indoor way finding system is implemented for blind people to independently access unfamiliar buildings. An unsighted person can discover vivid rooms and building exit or an escalator. This system includes text recognition. To discriminate between an office doors from a bathroom door, it extracts an discover the text information.

T. Rubesh Kumar etal [2] proposedreading is obviously essential in today's society. Printed text is everywhere in the form of reports, receipts, bank statements. There are already a few systems that have some promise for portable use, but they cannot handle product labeling. But a big limitation is that it is very hard for blind users to find the position of the bar code and to correctly point the bar code reader at the bar code [1]. T.Rubesh Kumar, C. Purnima have proposed a camera-based assistive text reading framework to help blind persons read text labels and product packaging from hand-held objects in their daily lives

In [3] Asif Shahbaad, Faisal Shaft, Andreas Dingell "ICVAR 2011 robust reading competition challenge 2: reading text in scene images"2011 international conference on document analysis and recognition of text in natural scene images is becoming a prominent research area because imaging devices like mobile phones are available. The ICDAR2011 robust reading event objection was formulated to access the act of novel algorithms in perceive and disclose content from complicated images.

System analysis:

Problems with Existing System

Extraction of text regions in natural images, the algorithm is robust in most cases, except for very small text characters that are not properly detected. Only if the text is in large or medium size this technology works properly, if the text is smaller in size it does not detect text accurately.

In Tyflos, the speech user integration might not work perfectly in a noisy environment, rendering it limited to indoor use. If the speech user integration does not work properly the user faces difficulty for accessing printed text.

Finger Reader is one such device, a wearable ring with a camera which is presented on the front. The voice user interface might not function perfectly in a chaos surrounding, rendering is restricted to indoor needs.

"Extraction of text regions in Natural images," the algorithm is robust in most common cases, but doesn't detect text, when it is in small size.

Proposed System

The proposed System helps the user to read the printed text efficiently.

The System first scans the text and image is captured through an application known as IP webcam.

IP web cam is an application that converts android device into internet camera with multiple view options that can be seen on any platform using VLC player or an internet browser. The video can be viewed via a browser or streamed over phone.

After the image is captured, text extraction is performed on the image.

The text extracted from the image is converted to audio which is available to the user.

Advantages of Proposed System

Text detection and voice output which helps visually impaired to read text with ease.

Maintains sequential order and does not require any positioning constraint for reading text.

Uses less hard ware.

This system provides accurate results with all font sizes and styles.

Feasibility Study

Technical feasibility

Does the necessary technology exist to do what is suggested? The open CV library of python provides the necessary pytessaract text extraction module and tkinter library for user interface.

Can the system upgraded if developed? - The project has been implemented using basic python open CV but can be improved using tensorflow, a deep learning frame work supported by open CV.

Operational feasibility:

User friendly

Portability

Availability

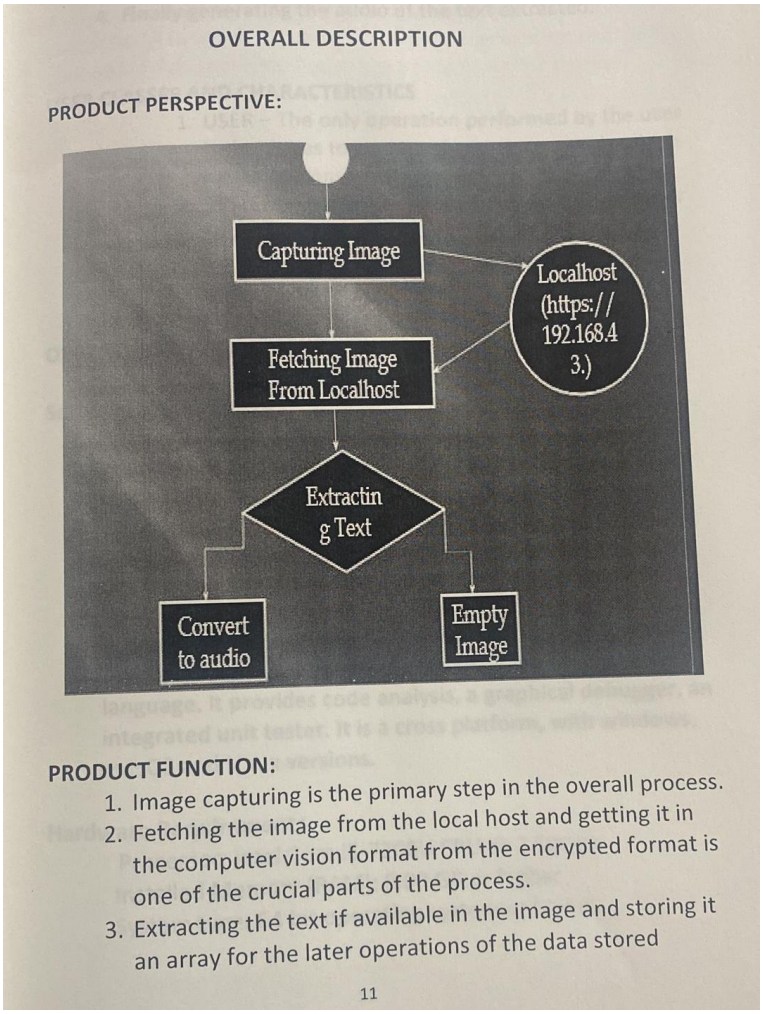


Fig 1

Product function

1. Image capturing is the primary step in the overall process.
2. Fetching the image from the local host and getting it in the computer vision format from the encrypted format is one of the crucial parts of the process.
3. Extracting the text if available in the image and storing it an array for the later operations of the data stored
4. Finally generating the audio of the text extracted.

User classes and characteristics

1. **User:** The only operation performed by the user is that it has to capture the image properly which has to be converted to audio.
2. **System:** It has 2 major tasks to be done in order to get the correct outcome. They are:
3. Image to text conversion
4. Audio generation

Operating Environment

Software Requirements

1. Programming language: Python
2. Windows Xp or higher
3. IP web cam
4. Tesseract

Open CV supports the deep learning frame works tensor flow.

Pycharm IDE: it is an integrated development environment used in computer programming, specifically for python language. It provides code analysis, a graphical debugger, an integrated unit tester. It is a cross platform, with windows, macOS and Linux versions.

Hardware Requirements

Processor: Intel Core 5-7200U CPU @ 2.50GHz

Installed Memory (RAM): 8.00 GB or higher

System type: 64 bit operating system or higher

A data flow diagram (DFD) maps out the flow of information for any process or system. It uses defined symbols like rectangles, circles and arrows, plus short text labels, to show data inputs, outputs storage points and the routes between

each destination. Data flowcharts from simple, even hand-drawn process overviews, to in-depth, multi-level DFDs that dig progressively deep.

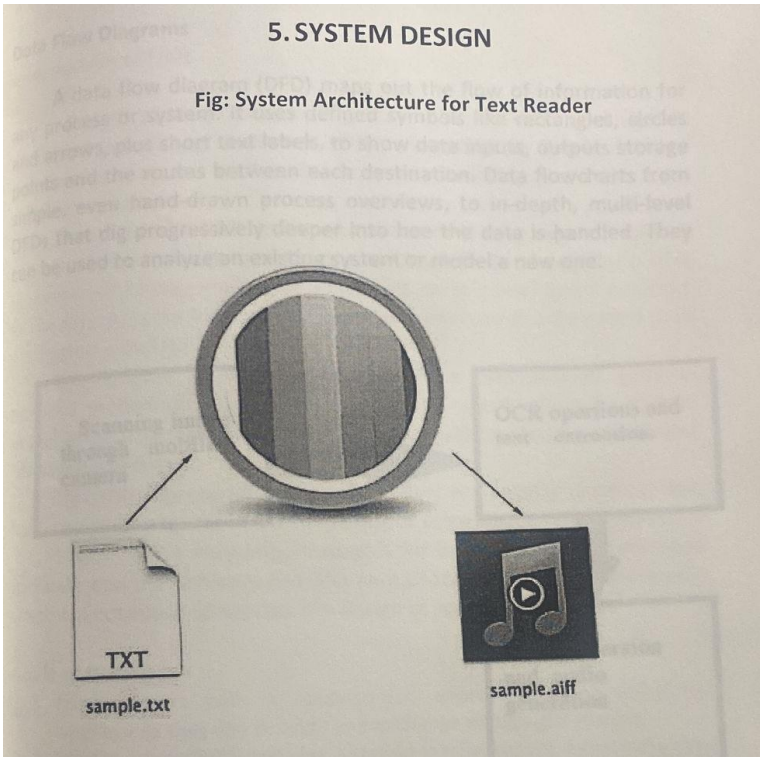


Fig 2: System architecture for text reader

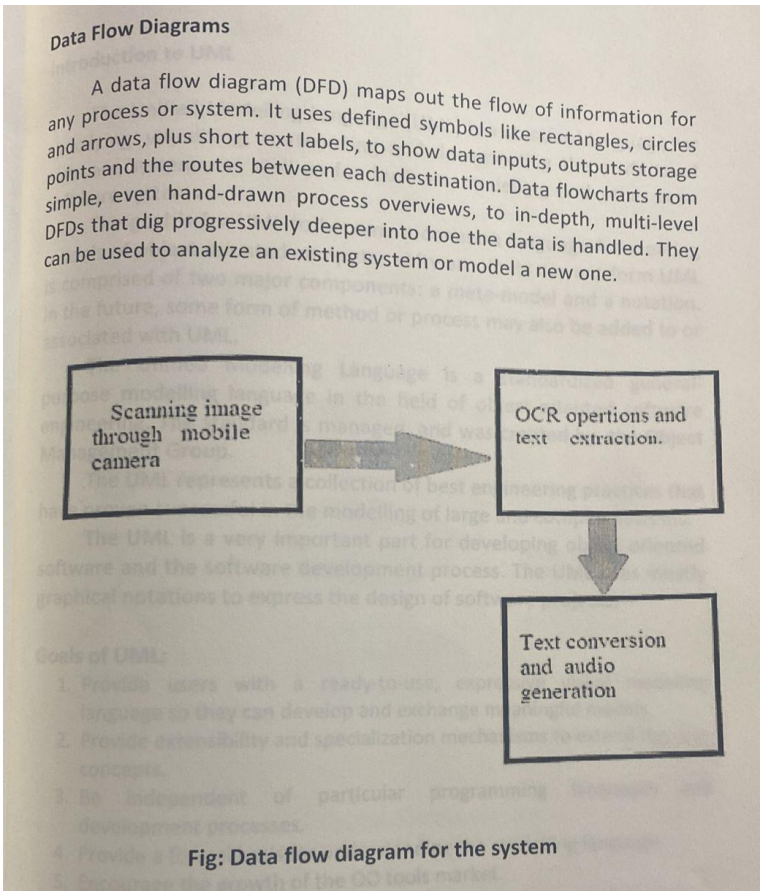


Fig 3: Data flow diagram for the system

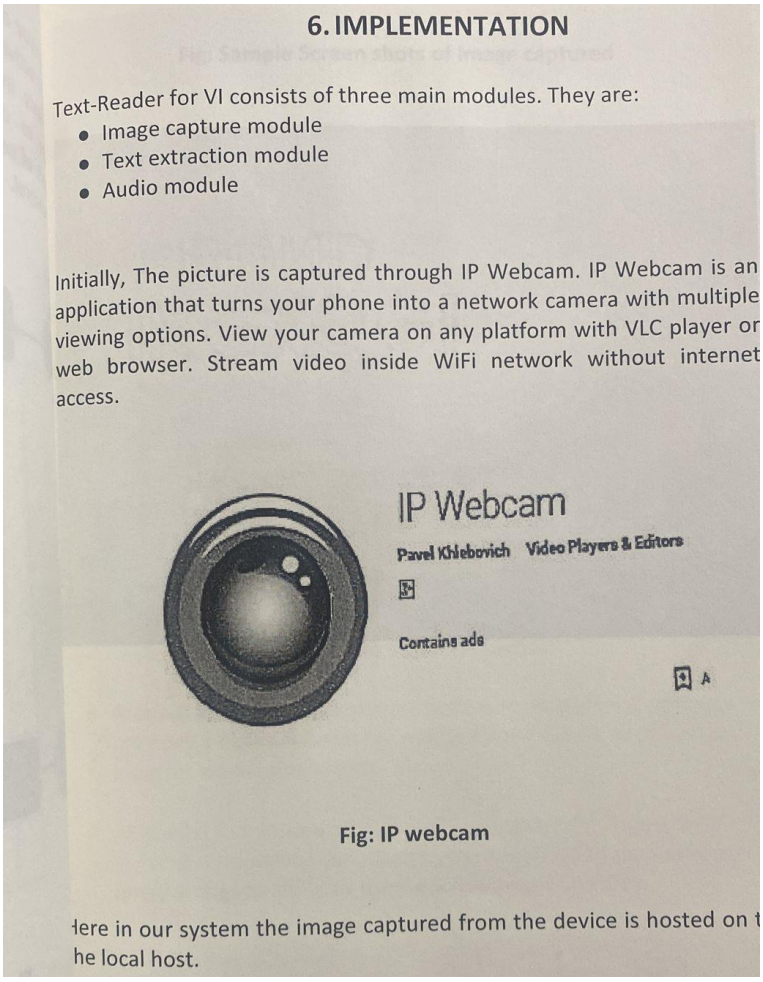


Fig 4: IP webcam

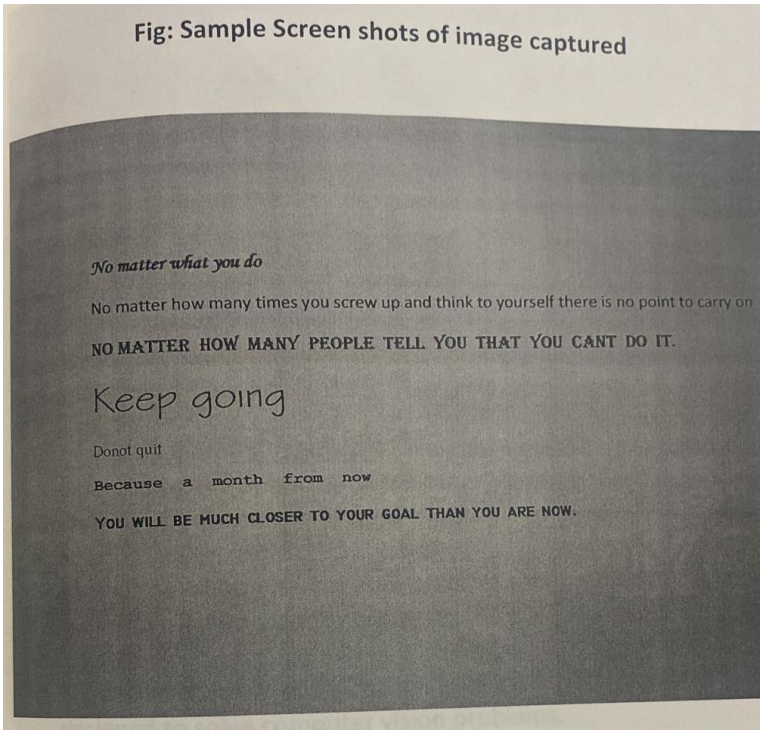


Fig 5: Sample screen shots of image captured

MODULE 1:

IMAGE CAPTURE MODULE:::

Image processing module is responsible for capturing the image pointed by camera. The interface used by the system and the capturing device is IP-webcam.

Code written to perform the above operations:

TextReader.py

```
import requests
import cv2
import numpy as np


url="http://192.168.0.3:8080/shot.jpg"
img_resp=requests.get(url)
#print(img_resp.status_code)
img_arr=np.array(bytearray(img_resp.content),    =np.uint8)
#print(img_arr)
img=cv2.imdecode(img_arr,-1)
cv2.waitKey(10)
```

Fig: Code for image received from the url

Fig 6: Code for image received from the URL

MODULE 2:

TEXT EXTRACTION MODULE:

The text extraction module is responsible for recognizing and extracting the text from the image.
It is done by using **pytesseract**-an **OCR** tool for python.
Optical Character Recognition (OCR) is the process of electronically extracting text from images or any documents like PDF.

Pytesseract:

- Python-tesseract is an optical character recognition (OCR) tool for python.
- It will recognize and "read" the text embedded in images.
- Python-tesseract is a wrapper for Google's Tesseract-OCR Engine.
- It is also useful as a stand-alone invocation script to tesseract, as it can read all image types supported by the Python Imaging Library.
- Whereas tesseract-OCR by default only supports tiff and bmp.

The system take the CV formatted image and gives it as an input to the pytesseract.
The py-tesseract uses its image_to string () method to extract the text and stores it in a variable. By using the strip () method, all the unnecessary whitespaces is deducted from the obtained text so as avoid the garbage values interception.

Fig 7

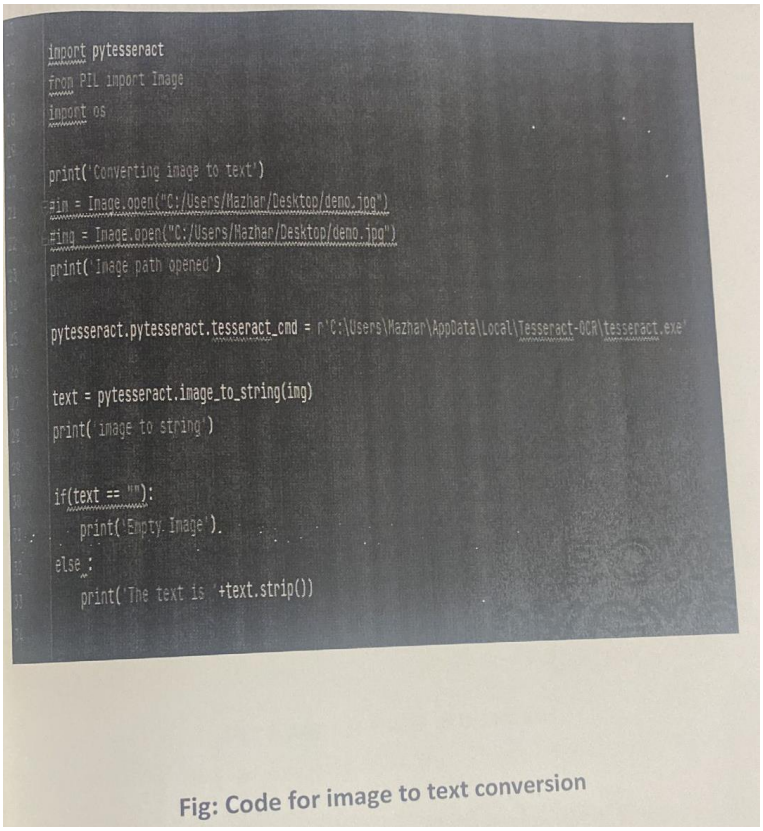
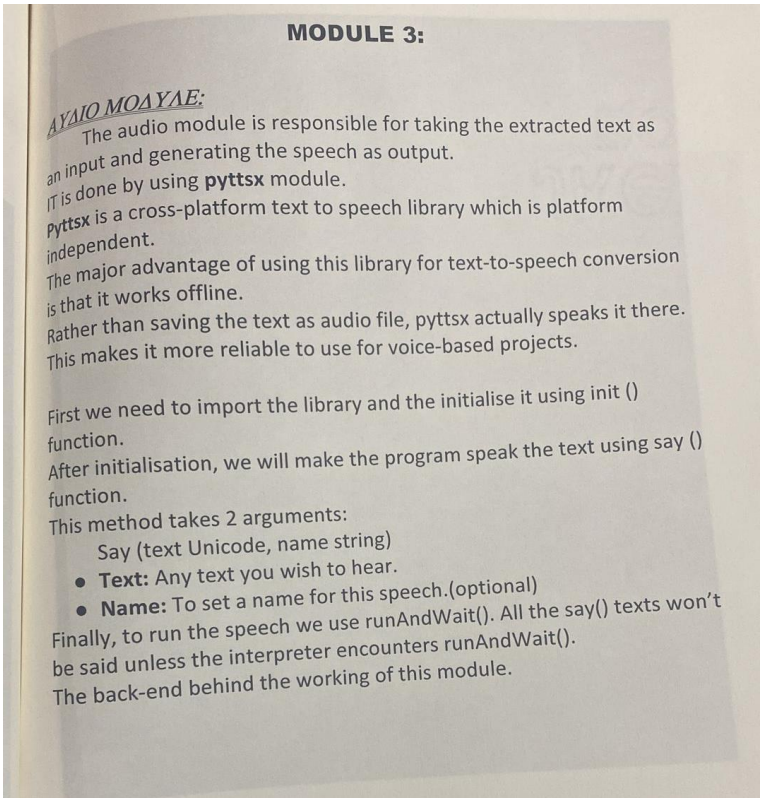


Fig 8: Code for image to text conversion



MODULE 3:

ΑΥΔΙΟ ΜΟΔΥΛΕ:

The audio module is responsible for taking the extracted text as an input and generating the speech as output.
IT is done by using **pyttsx** module.
Pyttsx is a cross-platform text to speech library which is platform independent.
The major advantage of using this library for text-to-speech conversion is that it works offline.
Rather than saving the text as audio file, pyttsx actually speaks it there. This makes it more reliable to use for voice-based projects.

First we need to import the library and the initialise it using init () function.
After initialisation, we will make the program speak the text using say () function.
This method takes 2 arguments:

Say (text Unicode, name string)

- **Text:** Any text you wish to hear.
- **Name:** To set a name for this speech.(optional)

Finally, to run the speech we use runAndWait(). All the say() texts won't be said unless the interpreter encounters runAndWait().
The back-end behind the working of this module.

Fig 9

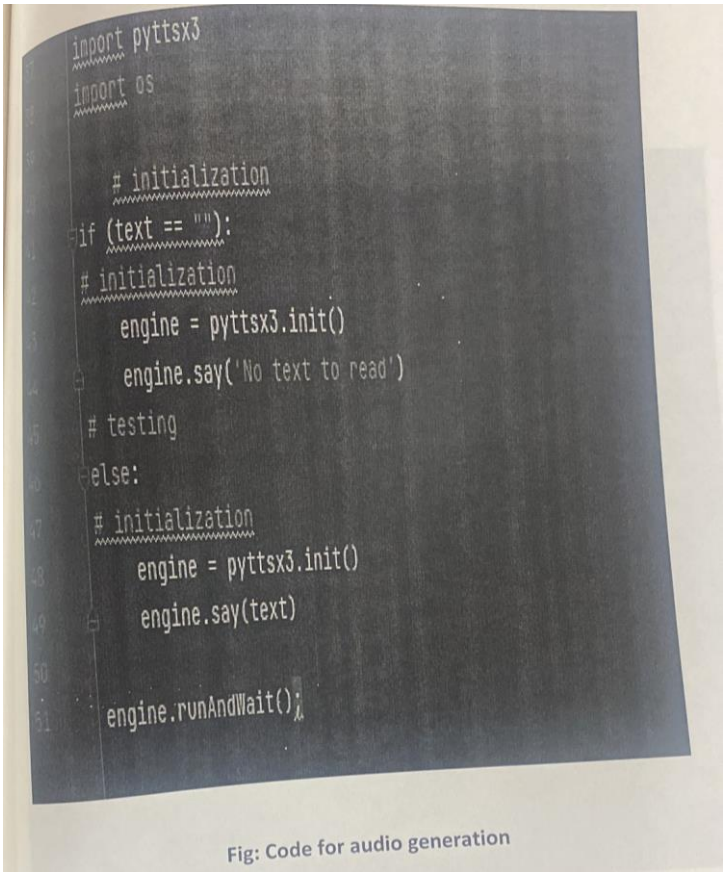


Fig 10: Code for audio generation

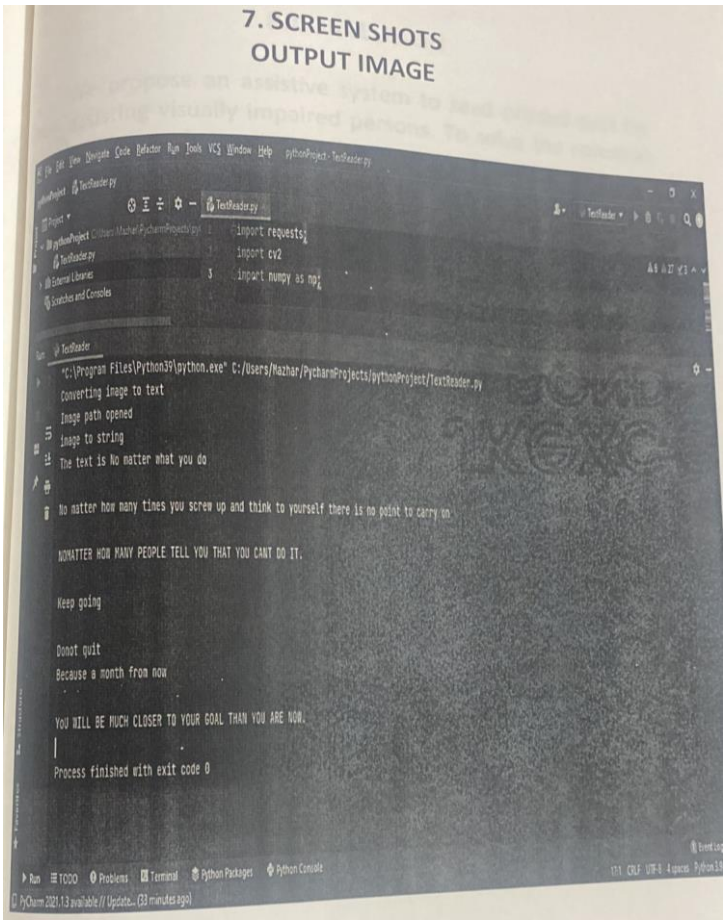
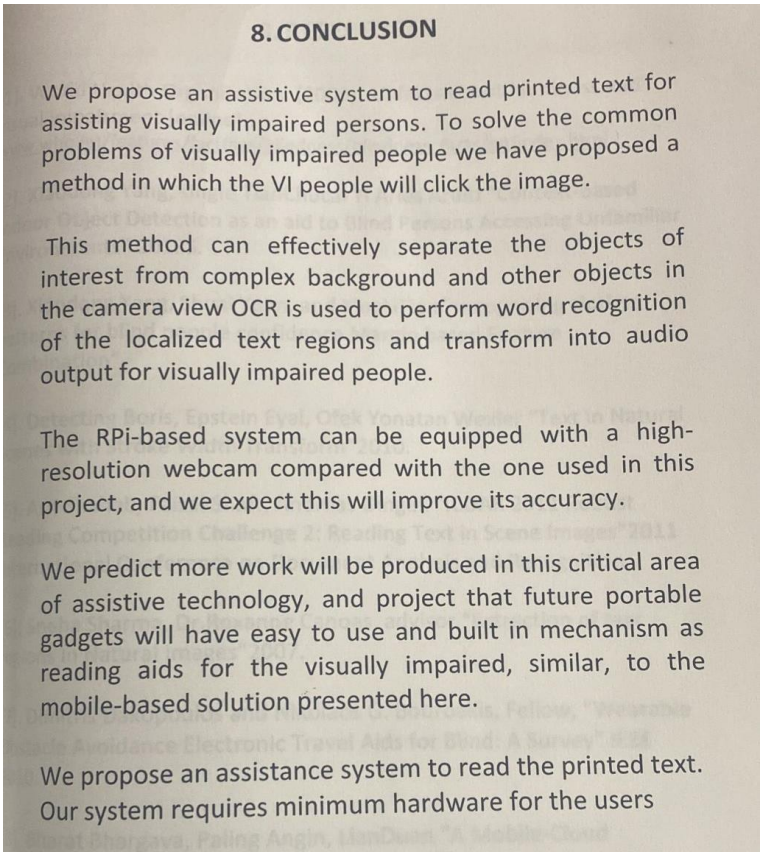


Fig 11


8. CONCLUSION

We propose an assistive system to read printed text for assisting visually impaired persons. To solve the common problems of visually impaired people we have proposed a method in which the VI people will click the image.

This method can effectively separate the objects of interest from complex background and other objects in the camera view OCR is used to perform word recognition of the localized text regions and transform into audio output for visually impaired people.

The RPi-based system can be equipped with a high-resolution webcam compared with the one used in this project, and we expect this will improve its accuracy.

We predict more work will be produced in this critical area of assistive technology, and project that future portable gadgets will have easy to use and built in mechanism as reading aids for the visually impaired, similar, to the mobile-based solution presented here.

We propose an assistance system to read the printed text. Our system requires minimum hardware for the users

Fig 12

The audio module is responsible for taking the extracted text as an input and generating the speech as output.
IT is done by using pytts module.
Pyttsx is a cross-platform text to speech library which is platform independent.
The major advantage of using this library for text-to-speech conversion is that it works offline.
Rather than saving the text as audio file, pytts actually speaks it there.
This makes it more reliable to use for voice-based projects.
First we need to import the library and the initialise it using init () function.
After initialisation, we will make the program speak the text using say () function.

This method takes 2 arguments
Say (text Unicode, name string)

- Text: Any text you wish to hear.
- Name: To set a name for this speech. (optional)

Finally, to run the speech we use runAndWait(). All the say) texts won't be said unless the interpreter encounters runAndWait().
The back-end behind the working of this module.

Conclusion
We propose an assistive system to read printed text for assisting visually impaired persons. To solve the common problems of visually impaired people we have proposed a method in which the VI people will click the image.
This method can effectively separate the objects of interest from complex background and other objects in the camera view OCR is used to perform word recognition of the localized text regions and transform into audio output for visually impaired people.
The RPi-based system can be equipped with a high-resolution webcam compared with the one used in this project, and we expect this will improve its accuracy.
We predict more work will be produced in this critical area of assistive technology, and project that future portable gadgets will have easy to use and built in mechanism as reading aids for the visually impaired, similar, to the mobile-based solution presented here.
We propose an assistance system to read the printed text.
Our system requires minimum hardware for the users

References

1. World Health Organisation. 10 facts about blindness and visual impairment [online], 2009 www.who.int/features/factfiles/blindness/blindness_facts/en/index.html.
2. Xiaodong Yang, single TianChucai Vi Aries Arditi Context-based Indoor Object Detection as an aid to Blind Persons Accessing Unfamiliar Environments, 20105.
3. Xiaodong Yang, Shuai Yuan, and YingLiTian" recognizing clothes patterns for blind people confidence Margin based Feature Combination"
4. Detecting Boris, Epstein Eyal, Ofek Yonatan Wexler Text in Natural Scenes with Stroke Width Transform, 2010.
5. Asif Shahab, Faisal Shaft, Andreas Dingell "ICDAR 2011 Robust Reading Competition Challenge 2: Reading Text in Scene Images"2011 International Conference on Document Analysis and Recognition.
6. Sneha Sharma, Dr.Roxanne Canoas, advisor "Extraction of text regions in Natural Images, 2007.
7. Dimitris Dakopoulos and Nikolaos G. Bourbakis, Fellow, "Wearable Obstacle Avoidance Electronic

Travel Aids for Blind: A Survey" IEEE, 2010.
8. Bharat Bhargava, Paling Angin, LianDuan "A Mobile-Cloud Pedestrain Crossing Guide for the Blind".
9. Bingil Tian, Chula Vi "Assistive Text Reading from Complex Background for Blind Persons.