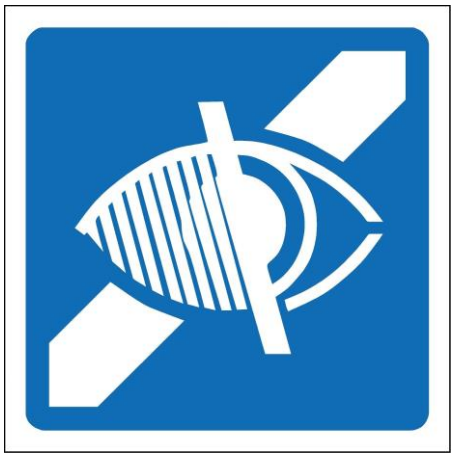
Credit Hours System

CCEN480 – GP1

Cairo University

Faculty of Engineering



Pocket Lens

GP 1 REPORT

Ahmed Mohamed Ismail	1180501
Mostafa Ashraf Ahmed	1180406
Moaz Mohamed Mohamed	1180528
Nader Youhanna Khalil	1180477

Supervised by

Assoc. Prof. Mona Farouk

Submitted On: 31/12/2022

Abstract

Even though many mobile devices today include accessibility features available for visually impaired and blind (VIB) users, many of these users are reluctant to use them. This is because the user interface is often designed for sighted people. This is caused by the fact that the main input and output methods in mobile devices are tactile or visual in nature. However, in recent years, there have been many innovative applications that assist VIB users in navigating their environment. Programmers have made use of technological advances regarding gyroscope sensors and vibration feedback to make communication possible. The proposed system relies on input images and videos provided by the user's device camera to allow daily life navigation without the need to use such sensors. It makes communication between VIB users and their devices possible using speech/text conversion techniques.

Table of Contents

1. Introduction 1
   1.1. Motivation and Justification 1
   1.2. Problem Statement 1
   1.3. Proposed Solution 1
2. Market Feasibility Study 2
   2.1. Market Survey 2
      2.1.1. Survey 2
      2.1.2. Competitors 2
   2.2. Feasibility Study 3
      2.2.1. Market Analysis 3
         2.2.1.1. Market Outline 3
         2.2.1.2. Competitive Study 4
         2.2.1.3. Market Research 4
         2.2.1.4. Cases in Egypt 5
         2.2.1.5. SWOT Analysis 5
      2.2.2. Technical Analysis 6
         2.2.2.1. Engineering and Technology 6
3. System Design and Architecture 7
   3.1. System Description 7
   3.2. System Block Diagram 8
   3.3. Modules 8
      3.3.1. Main Modules 8
         3.3.1.1. Virtual Assistant (Chatbot): - Ready Made 8
         3.3.1.2. Scene Descriptor: - From Scratch (Neural Networks) 10
         3.3.1.3. People and emotion recognizer: - OpenCV 11
         3.3.1.4. Text Reader: - Ready Made 12
         3.3.1.5. Product Identifier: - Ready made 13
         3.3.1.6. Currency Detector: - From Scratch 14
         3.3.1.7. Clothes Descriptor: - From Scratch 15
      3.3.2. Secondary Modules 16
         3.3.2.1. Note Taker: - Ready made 16

3.3.2.2. Event Scheduler: - Ready made 16
4. Time Plan 16
5. Task Division 17
6. Modules implementation 18
7. Conclusion 18
8. References 19

List of Figures

Figure 1: Market Research 4
Figure 2: Proposed Wearable 7
Figure 3: System Block Diagram 8
Figure 4: Mobile Voice Assistant Block Diagram 9
Figure 5: Scene Descriptor and Navigator Block Diagram 10
Figure 6: People Recognition Block Diagram 11
Figure 7: Text Reader Block Diagram 12
Figure 8: Product Identifier Block Diagram 13
Figure 9: Currency Detector Block Diagram 14
Figure 10: Clothes Detector Block Diagram 15

GRADUATION PROJECT REPORT

1. Introduction

1.1.Motivation and Justification

VIB users are often put at a disadvantage regarding their visually able peers. Technological advancements have always been concerned with providing better and easier to use solutions. These efforts have been largely directed toward the use of sensors, which can in many cases be unreliable.

Moreover, many of the applications that can be found in the market are not particularly easy to use. They often require some degree of tactile interaction, which VIB users will most probably not be able to provide. Some of these applications are designed to be used by sighted people alongside VIB users, which can come as impractical.

The before mentioned reasons led us to consider using AI and Machine Learning techniques to create a mobile application that can serve as an assistant to VIB people. We will be addressing these previous problems by rendering the contact between the application and the VIB user purely vocal as much as the desired features allow for it. In other words, the user will communicate with the chatbot through speech.

1.2.Problem Statement

Our objective is to provide VIB users with a system that sufficiently fulfills their needs in daily life navigation. The system should provide all the functionalities needed for daily life tasks with ease of use.

1.3.Proposed Solution

Our application Pocket Lens acts as a tool that can be used by VIB people in their daily life for navigation. Therefore, our goal is to facilitate daily life tasks which include identifying currency, clothes, people, products in markets, and other elements that may be in a scene and which prove essential for proper navigation. Our product also presents a text reader functionality that allows VIB users to perform the indispensable task of reading. The interface is proposed to convert speech to text and vice versa, to offer VIB users an application that can be used without the need for assistance.

2. Market Feasibility Study

2.1.Market Survey

2.1.1. Survey

We need to properly distribute the survey among many segments concerning their educational, financial, societal levels, age, and gender. This is crucial to ending up with a survey that represents the population accurately. This would need ample time and resources and since our access to visually impaired people would be limited, it will be difficult to collect enough data so we decided to collect data from surveys on the internet.

2.1.2. Competitors

A new set of smart glasses from Envision can help blind people read. The glasses, which were built on the enterprise edition of Google Glass, rely heavily on artificial intelligence to help people see and understand their surroundings.

Envision isn't the only company developing apps and hardware to help blind people see. Google's Lookout app can help people identify food labels, as well as find objects in a room. Like Envision, it can also scan documents, money, and other things.
As CNET mentions, Be My Eyes is also another app that can help. It connects visually impaired users with sighted volunteers. The volunteers then help the user get around via a live chat function.

Microsoft Soundscape is an application built by the Enable Group in Microsoft Research. The Soundscape app is breaking barriers and opening up new possibilities for visually impaired people with voice-based navigation. Anyone can take this app on the go and enjoy the independence that comes with being able to explore the world on their terms. Using a stereo headset such as Air Pods, users can traverse new and old environments guided by a map delivered in 3D sound.

There are also Facing emotions created by Huawei which identifies the 7 basic human emotions of irritation, contempt, sorrow, fear, anger, surprise, and happiness. The app then turns those emotions into unique sounds to help the visually impaired learn how the person on the other side of the conversation is feeling.

2.2.Feasibility Study

2.2.1. Market Analysis

2.2.1.1. Market Outline

Our Market	Mobile applications for helping people with visual loss in their daily lives such as shopping, dressing, or moving around the city
Target Audience	People with visual impairment
Competitors	• Google (lookout) • Envision • Microsoft soundscape • Facing Emotions
Customer Needs	• Navigate their homes easily • Chat with people and know their emotions • Identify the currency or amount they are holding • Pick matching clothes easily
Services Offered by Competitors	• Navigation • Scan documents, money • Identify food labels
Services Offered to Customers	• Chatbot • Navigator • Document scan • Read handwritten articles and documents • Emotion detector

2.2.1.2. Competitive Study

Features \ Apps	Lookout	Envision	Soundscape	Facing Emotions
Chatbot				
Scene Descriptor	Done		Done	
Text reader	Done	Done		
Emotion detector		Done		Done
Currency detector	Done			
Products detector	Done			

2.2.1.3. Market Research

To understand the market clearly, we must systematically gather and analyze the needed data. As shown in the below figure, visual impairment is a problem that faces approximately 36 million people.

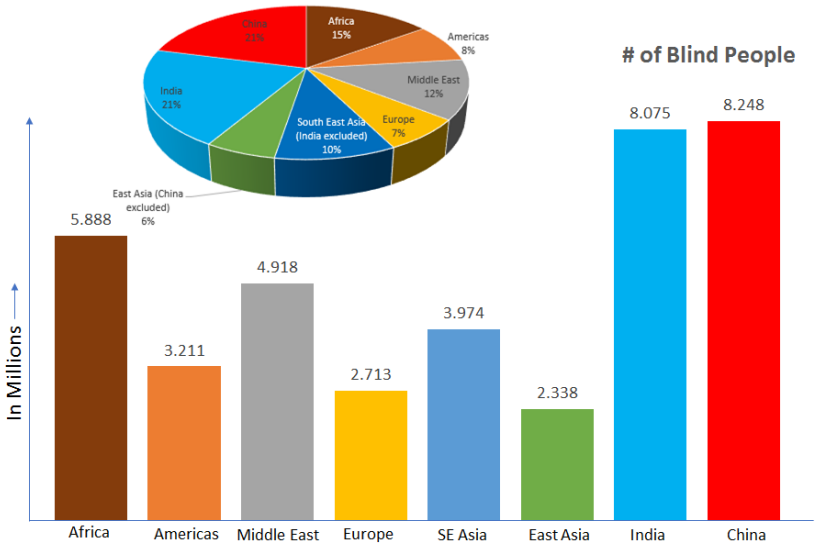


Figure 1: Market Research

- The number of blind people across the world is set to triple within the next four decades, researchers suggest.
- Writing in Lancet Global Health, they predict cases will rise from 36 million to 115 million by 2050 if treatment is not improved by better funding.
- While Analysis of data from 188 countries suggests there are more than 200 million people with moderate to severe vision impairment.
- That figure is expected to rise to more than 550 million by 2050.

2.2.1.4. Cases in Egypt

According to the World Health Organization W-H-O, there are more than 2.2 million people with visual impairment in Egypt 900,000 of whom are blind.

Khalid Karem, a graduate student in The Faculty of Languages at Ein Shams University specializing in English and Hebrew language who was born with full blindness said: "Inaccessibility of streets is another hardship that visually impaired people have to cope with. Sidewalks and pavements that are supposed to help the blind person walk safely are instead occupied by street sellers, shops, and parking cars. Visually disabled people who decide to walk alone in the streets of Cairo put their lives at risk. Karem, recalls his experience walking in Cairo."

2.2.1.5. SWOT Analysis

Strengths	Weaknesses
• AI chatbot • 24/7 availability • Privacy • Internet access not needed(maps can be saved offline) • Understand other people better by knowing their facial emoticons	• AI functions are not 100% accurate hence can identify wrong currencies or emotions • Lack of empathetic human care • Expertise need
Opportunities	**Threats**
• Improving humans life • Convenience • Spreading awareness	• Risk of over-depending on the technology • Data security

The SWOT analysis shows that we have great strengths that could contribute to great opportunities. However, we must be careful to avoid probable threats and overcome our weaknesses. The several technologies offered by our application like AI chatbot will not only spread awareness but also it will compensate for empathetic human care. Also, offering privacy as a key feature in our application will contribute to anonymity and overcome data security concerns. Moreover, the idea of being able to use the features 24/7 will help make the u to be more independent in their lives and avoid the feelings of always needing to depend on others to perform daily life tasks

2.2.2. Technical Analysis

Technical analysis is one of the most important dimensions of the feasibility study. We will analyze the project to see if it is technically feasible or not and to make the best out of the available technology and human resources or decide on an alternative. There are a lot of factors that we will discuss.

2.2.2.1. Engineering and Technology

- document scan → NLP
- Chat Bot → NLP
- Mobile application → Flutter

3. System Design and Architecture

3.1.System Description

Our application targets visually impaired people, its focus is to help them in their everyday lives. It is a mobile app and since the users are practically blind the inputs to our system are voice and images captured from the mobile microphone and camera respectively. The proposed way to use the app is shown in Figure(2) which is a wearable mobility assistive system by using an Android smartphone and a sling pouch.

Figure 2: Proposed Wearable

The main interface provided to the user is a virtual assistant (chatbot) that takes commands from him via speech and then redirects the user to the corresponding module to do the required task which eventually responds to the user in the form of speech feedback.

3.2.System Block Diagram

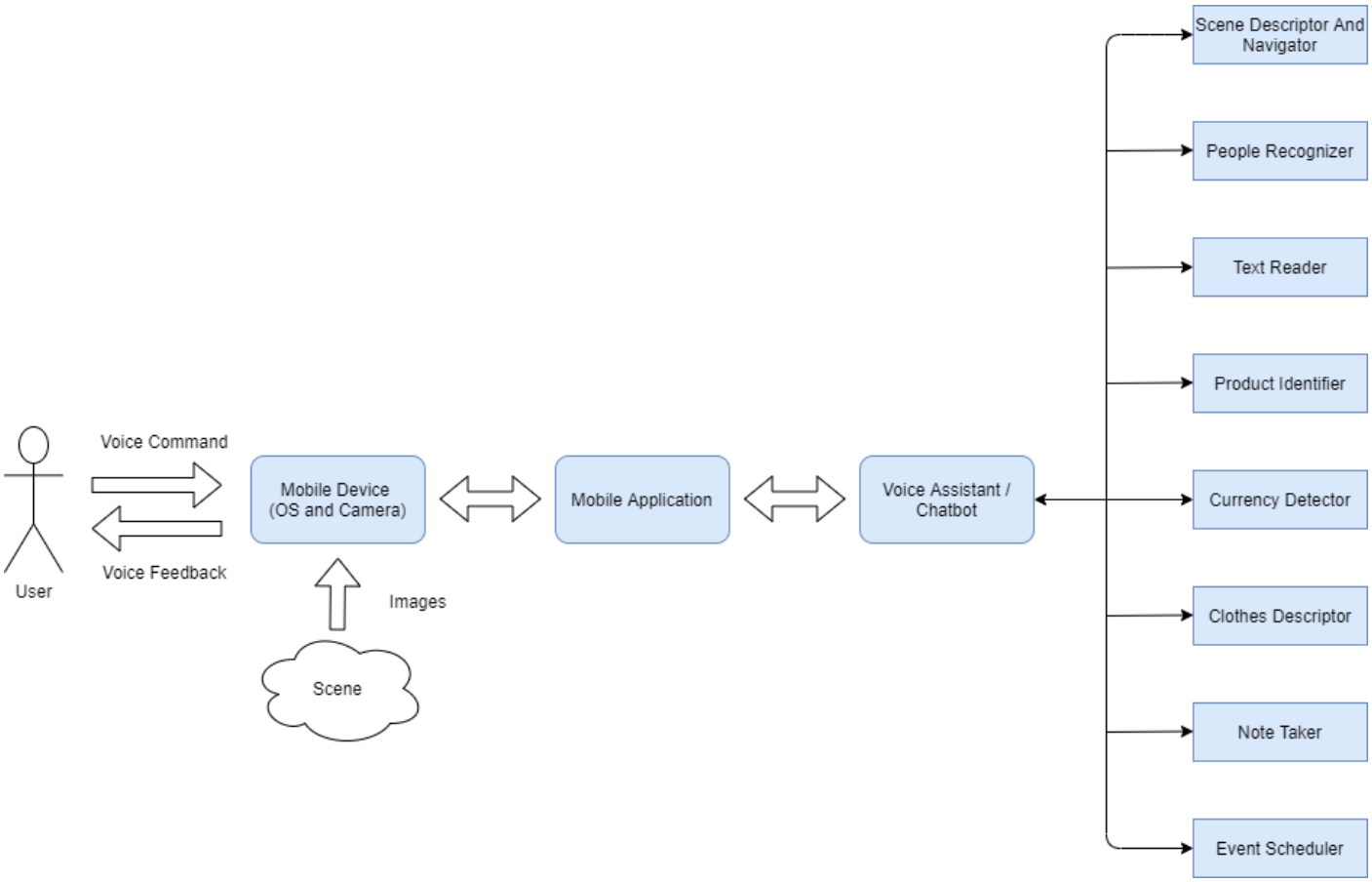


Figure 3: System Block Diagram

3.3.Modules

3.3.1. Main Modules

3.3.1.1. Virtual Assistant (Chatbot): - Ready Made

This module serves as an assistant that carries our everyday tasks via voice command. It recognizes the user's voice and accomplishes the task requested from it.

Our application helps the user to search for things on the browser, set alarms, ask for the weather, and do any other everyday task. Also, it redirects the user to the other modules he requests, which are listed below.

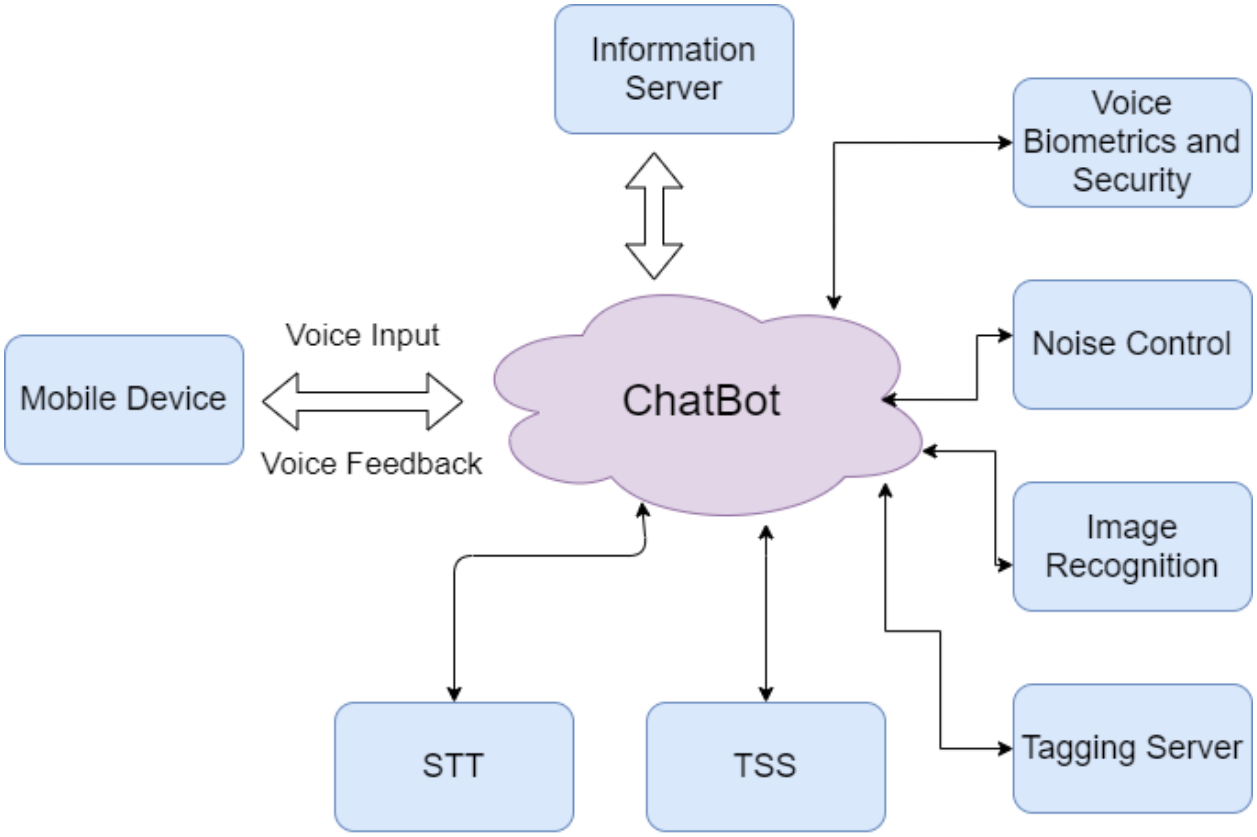


Figure 4: Mobile Voice Assistant Block Diagram

STT: Speech to Text, this is the process of converting speech signal into digital/text data.

TTS: Text to Speech, this is the process of converting digital/text data into speech.

Intelligent tagging and decision-making serve for interpreting the user's request.

Image recognition: It is used for identifying places, and people within images.

Noise control: reduce and eliminate the background noise for voice clarity.

Voice biometrics and Security: Identity who is talking to it (customized assistant to your voice for security).

Speech compression: resize the voice data and send it to the information server in a succinct format.

Voice interface/feedback: the response that the user receives as feedback for his request.

However, there are ready-Made solutions for voice assistants like Siri, Google Assistant, Alexa, and Cortana so this module will not be implemented from scratch, it will be customized for our app purpose.

3.3.1.2. Scene Descriptor: - From Scratch (Neural Networks)

This module describes the scene that the camera captures in from of the user so that he can imagine what the scene in front of him looks like.

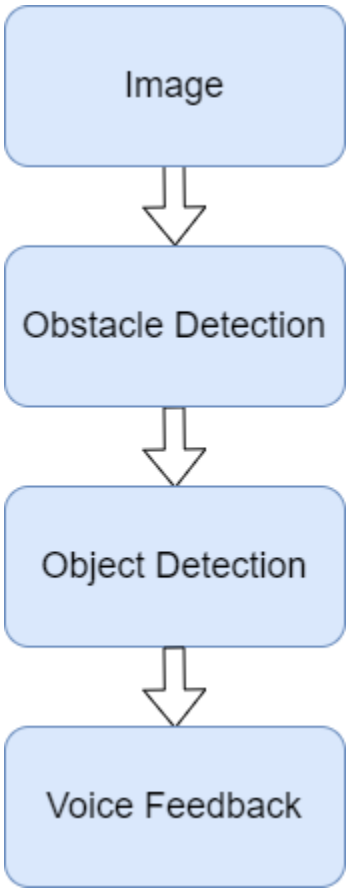


Figure 5: Scene Descriptor and Navigator Block Diagram

- We will use some ready-made libraries.
- Object detection: YOLO algorithm, CNN
- Obstacle detection: YOLO algorithm, CNN
- Calculate Distance: ARCore Depth Lab API by Google

3.3.1.3. People and emotion recognizer: - OpenCV

This module is responsible for recognizing the people from the set of the user's friends and family and notifying the user of their presence. Also, it recognizes their emotions from their facial expressions.

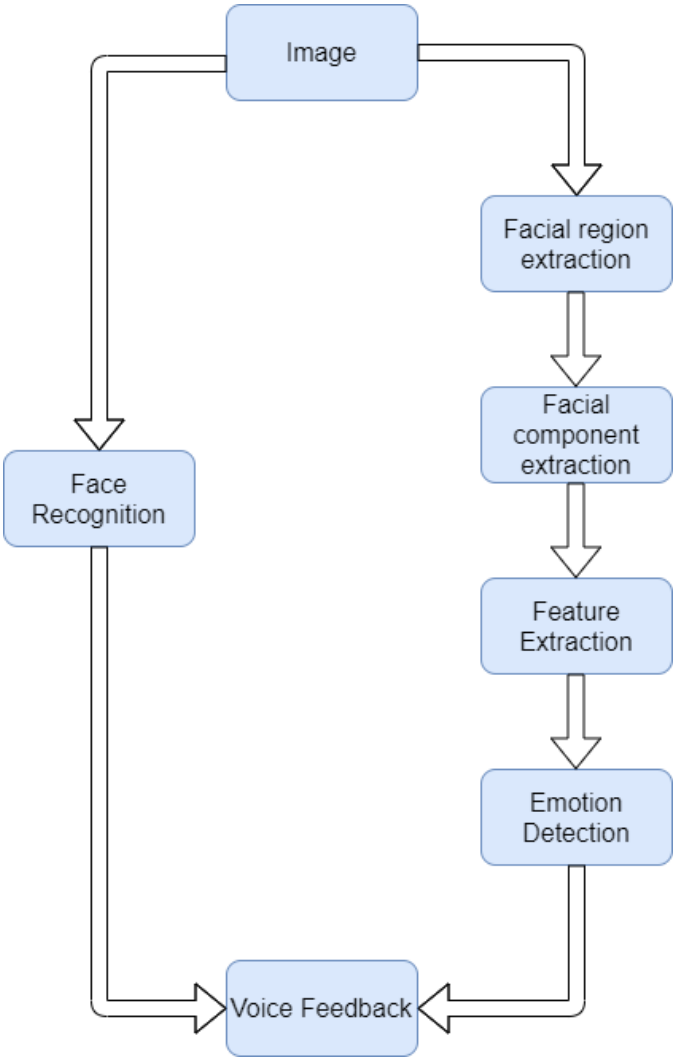


Figure 6: People Recognition Block Diagram

Face Detection: Viola Jones Algorithm

Face Recognition: DeepFace, DeepID, series of systems, VGGFace, FaceNet, FischerFaces/EigenFaces

Emotion Detection: Histogram segmentation, Feature extraction, fuzzy classifier-based emotion detection

3.3.1.4. Text Reader: - Ready Made

This module provides audio guidance to capture a printed page, recognizes the text and its original formatting, and then speaks the document. We will be using CNN to implement this module.

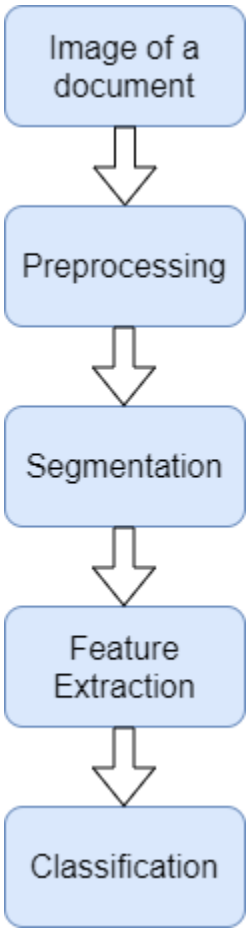


Figure 7: Text Reader Block Diagram

Preprocessing consists of

1. Noise removal
2. Binarization. i.e., converts the image into a binary image
3. Morphological operation. This will increase or decrease the size of the image to a constant image size.

3.3.1.5. Product Identifier: - Ready-made

The product identifier module is very useful for visually impaired people. This module takes an image containing a product and its barcode as an input and outputs voice feedback telling the user what product he just captured. For the user to find the barcode, they will have to keep rotating the object slowly in front of the camera. When the module sees a barcode, it will start beeping. This is very helpful for shopping.

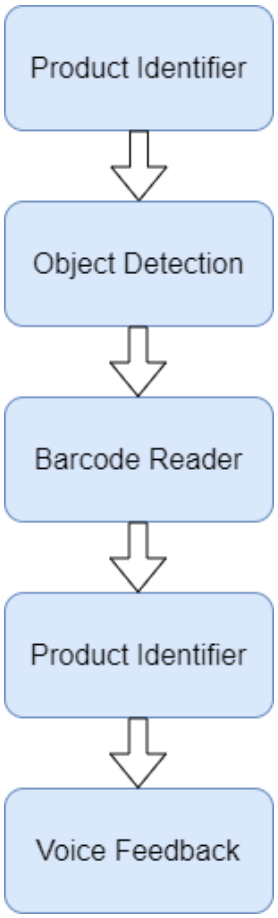


Figure 8: Product Identifier Block Diagram

3.3.1.6. Currency Detector: - From Scratch

This module will help visually impaired people recognize currencies. It will also help them count the currencies. This module has five steps, preprocessing techniques for removing noises and preparing the image for the next operations, segmentation, and ROI extraction processes in the second and third steps for extracting the foreground currency from the background, applying ORB Algorithm in step four, and finally matching the results with the dataset.

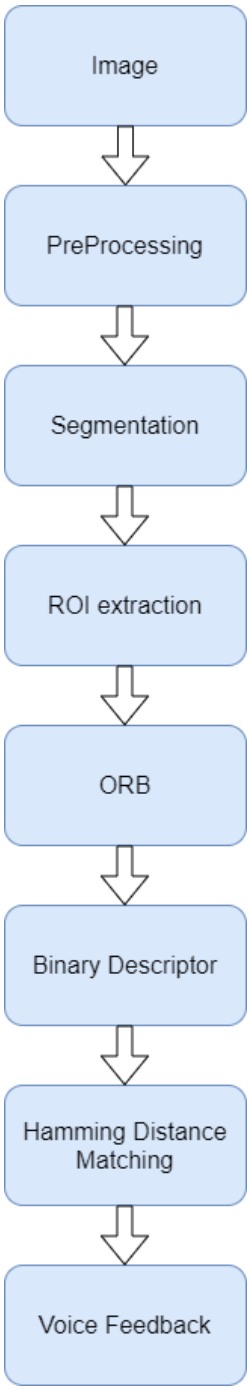


Figure 9: Currency Detector Block Diagram

3.3.1.7. Clothes Descriptor: - From Scratch

This module will be implemented from scratch. Its input consists of an image or a sequence of images. The output consists of whether clothes are to be found in the scene, and the type, color, and texture if yes.

Clothes detection and extraction: Morphological operations + Segmentation techniques + CNN

Texture Analysis: LBPH algorithm + GLCM Features + Law's Texture Energy Features

Classifier: Final classifier (SVM)

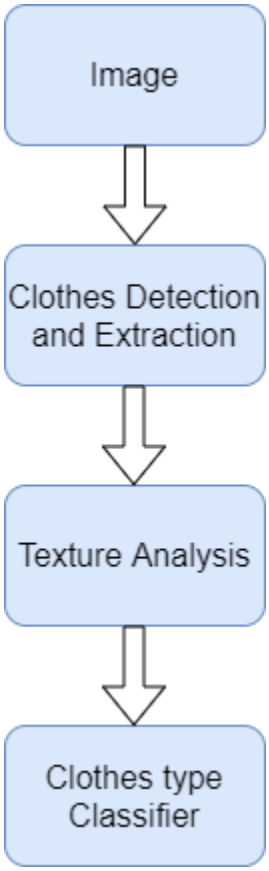


Figure 10: Clothes Detector Block Diagram

3.3.2. Secondary Modules

3.3.2.1. Note Taker: - Ready-made

3.3.2.2. Event Scheduler: - Ready-made

These two modules will be part of the existing application on the user's mobile device. The chatbot will redirect the user to the corresponding application, convert speech to text, and then set the event or register the requested note.

4. Time Plan

Week	Tasks	
1	AI Model Implementation	User Interface Design
2	Data Collection	Mobile Application development
3	Data Preprocessing	
4	AI Model Training	Implementation of non-AI modules
5	AI Model Performance Analysis	
6	AI Model Refining	Integrate Modules with Mobile Application
7	Integrate AI Modules with Mobile Application	
8	-	
9	Testing and Enhancement	
10		
11		
12		
13	Report Assembly	
14	Project Submission and Evaluation	

5. Task Division

Student	Task
Ahmed Mohamed	• Product Identifier • Currency Detection • Scene Descriptor
Mostafa Ashraf	• People recognizer • Emotions detector • Chatbot • Scene Descriptor
Moaz Mohamed	• Text Reader • Event Scheduler • Mobile App Development • Scene Descriptor
Nader Youhanna	• Clothes Detection • Note Taker • Mobile App Development • Scene Descriptor

6. Modules implementation

Module	Implementation
Virtual Assistant	Ready Made
Scene Descriptor	From Scratch (Neural Networks)
People and Emotion recognizer	OpenCV
Text Reader	Ready Made
Product Identifier	Ready Made
Currency Detector	From Scratch
Clothes Descriptor	From Scratch
Note Taker	Ready Made
Event Scheduler	Ready Made

7. Conclusion

In conclusion, our application has the goal of helping VIB users in their daily life tasks by enabling them to perform basic functions that are particularly difficult for them. Our product aims to use images and video footage provided by the user's device camera in identifying scene elements and people that are necessary to know and locate for proper navigation. The product also aims to be used completely by VIB users without the need for assistance from sighted people.

8. References

- https://www.cleveroad.com/blog/how-to-create-virtual-assistant-apps-like-siri-and-google-assistant/#:~:text=How%20to%20make%20a%20virtual,like%20Wit.ai%20or%20Jasper.
- https://www.inclusivecitymaker.com/apps-blind-visually-impaired-people/
- https://www.everydaysight.com/best-apps-for-visually-impaired/
- https://web.archive.org/web/20220506043755id_/https:/dl.acm.org/doi/pdf/10.1145/3491101.3503814
- http://ijasret.com/VolumeArticles/FullTextPDF/831_36.VIRTUAL_ASSISTANT_FOR_BLIND_PEOPLE.pdf
- https://drive.google.com/drive/folders/1mQ-Gul-ri8XNKHMReG3P8dKI1zqSP_ev
- https://link.springer.com/content/pdf/10.1007/s10209-020-00733-8.pdf
- https://ijsret.com/wp-content/uploads/2021/03/IJSRET_V7_issue2_211.pdf
- https://pdfs.semanticscholar.org/2590/ccf9445b96ef6ec17de8adae603f420517e2.pdf
- https://koreascience.kr/article/CFKO200533239341594.pdf
- https://asset-pdf.scinapse.io/prod/1673002162/1673002162.pdf
- https://www.ripublication.com/ijaer18/ijaerv13n2_44.pdf
- https://link.springer.com/chapter/10.1007/978-3-642-13161-5_8
- https://doi.org/10.1007/978-981-16-6482-3_44
- https://www.academia.edu/37509222/A_SMART_MOBILE_APPS_FOR_BLIND_USER
- https://www.researchgate.net/publication/338363439_Smartphone_apps_for_visually_impaired_persons
- https://www.researchgate.net/publication/304651244_VOICE_RECOGNITION_SYSTEM_SPEECH-TO-TEXT

- https://turcomat.org/index.php/turkbilmat/article/download/12651/9133/22435
- https://core.ac.uk/download/pdf/327266589.pdf
- https://doi.org/10.1007/978-3-319-23231-7_43
- https://doi.org/10.1016/j.cviu.2015.11.017
- https://www.researchgate.net/publication/329487411_Currency_Recognition_System_for_Blind_people_using_ORB_Algorithm