Faculty Of Engineering, Helwan University

Computer and Systems engineering Department

Blind Smart Virtual Assistant

Presented by:

Aml Ashraf Zayed Mohamed

Rana Mahmoud Kamal Mahmoud

Rasha Saad Fawzy Mohammed

Supervisor

Dr. Rasha Fathy Aly

Individual contributions

Team member	Contributions
Aml Ashraf Zayed	Speech to text Coding of text to speech Coding of translate coding of face recognition currency recognition integrate classes together Class Diagram Testing Enumerated Nonfunctional Requirements Connecting camera with raspberry Connecting GPS with raspberry
Rasha Saad Fawzy	Coding of text to speech Coding of translate Color recognition technique of face recognition currency recognition send location integrate classes together sequence diagram Testing Enumerated functional Requirements Connecting camera with raspberry Connecting GPS with raspberry
Rana Mahmoud kamal	technique of text to speech technique of translate technique of OCR coding of face recognition currency recognition integrate classes together use case diagram Testing Functional Requirements Specification Connecting camera with raspberry Connecting GPS with raspberry

Abstract

with the emergence of virtual assistants and their increased popularity, the possibility of generating one of our own has raised, using neural network algorithms and artificial intelligence to recognize commands and execute the actions commanded through natural language, allowing accessibility and inclusion for people with visual disabilities.

Blind people find it more challenging to move about independently because of their compromised vision, Moreover, a blind person's capacity to navigate in a given setting, along with their ability to organize their daily activities is vital to their health and wellbeing. The more saddening fact is that there are tens of millions of visually impaired people worldwide who must go through such an experience and are dependent on others for their wellbeing and happiness. The encouraging news, however, is that the rapid advancement in technology has seen the innovation of better systems for assisting the disabled, including the blind, So "Blind Smart virtual assistant" is designed to help the blind people to read and translate the typed text which is written in the English and Arabic language, this module is provided with technology to scan any written text and convert it into audio text. Also, it can translate words from English to Arabic and vice vera using Google API and recognize People, Currency and Color and send his location for someone he selects when he feels that he is lost in somewhere.

Keywords:
Blind people, Raspberry pi, OCR (Optical Character Recognition), GTTS (google text to speech), Speech Recognition, Voice commands, GPS Module, headphones.

Acknowledgment

First and foremost, praises and thanks to Allah, the almighty, for his showers of blessings throughout our whole journey of finding a beneficial idea and working as a great team on the project.

We would like to express our deep and sincere gratitude to our project supervisor, Professor Rasha Fathy Aly, for her enthusiasm, patience, insightful comments, helpful information, practical advice, and unceasing ideas that have always helped us tremendously in our research.

It was a great privilege and honor to work and study under her guidance. We are extremely grateful for what she has offered us. We truly hope that our outcome can make her proud of us.

Table of Contents

Table of figures

1. Introduction

Vision, to begin with, is the most crucial part of human psychology, every five seconds a person in the world turns blind. And every minute a child in the world turns blind, The Central Agency for Public Mobilization and Statistics revealed, according to the latest census conducted by the state, that the number of people with disabilities reached 8.636 million of whom 6.608 million people have mild difficulty, 1.636 million people have great difficulty, and 390.9 thousand people have absolute difficulty.

As the number of people with very difficult vision reached 439.2 thousand In our lives, there are many people who are suffering from blindness and vision difficulties, most blind people are smart people and can study if they have the chance to be able to study in normal schools because there aren't government school everywhere so by "Blind smart virtual assistant", These Inventions consider a solution to motivate blind students to complete their education despite all their difficulties, the percentage of educated people will increase,

Figure 1 Blind Student

The project presented here is a device that will allow blind people and people with visual disabilities to be independent of a companion or to have knowledge of braille language.

Figure 2 Braille language

Among all assistive devices, wearable devices are found to be the most useful because they are hand free or require minimum use of hands not only that but also dealing with our module only through voice commands

1.1 Problem Statement

Visually Impaired people face a big problem to live normally as they always need an assistant to do many functions in their lives, they cannot do themselves

Figure3 Assistant with blind person

Blind People and people with vision difficulties can study as normal students if they have an appropriate chance. Most blind people and people with vision difficulties did not study and that is because special schools for people with special needs not everywhere and most of them are private and expensive or they study at home acquiring basic knowledge from their parents, also blind people find difficulty to recognize people around them and currency they own, most people thought blind people and people with vision difficulties cannot live alone and they need help all the times. In fact, they do not need help all the time, they can depend on themselves most of the time and they have the chance to live like a normal person in this life.

The main reason to implement "Blind Smart Virtual Assistant" for blind people and people with vision difficulties is to prove for all people that blind people and people with vision difficulties have the chance to live a normal life with normal people and study in any school or university without the need for help all the time. By "Blind Smart Virtual Assistant", the percentage of educated people will increase Most Schools will be able to accept people with vision difficulties instead of open special schools.

1.2 Related works

the “blind modules” were simple and provide basic tasks which serve as a front-end display for the remote system. Now “Smart Glasses” become more efficient and provide several features.

We built our project on the electronic devices for the Blind and Visually Impaired as a prior works like:

-Oton Glass

The glasses have two small cameras and an earpiece. The camera captures pictures of words that user wants to read and reads out the words for the user via the earpiece.

The smart glasses are designed to help dyslexic to read.

They look like normal glasses (1)

Drawbacks:

Oton Glasses will only help people who have difficulty reading but Blind people will not benefit from it.

-MyEye2

OrCam MyEye is suitable for all eye conditions and all levels of vision loss, as well as for people with reading difficulties. The device increases independence by allowing individuals to access visual information (text, faces, products, colors, money notes), conveyed by audio. It will not improve a person’s vision. Hearing impaired individuals would not be able to benefit from the device. (2)

Drawbacks:

These glasses are not affordable for everyone because they are expensive.

-Aira

Airahas built a service that basically puts a human assistant into a blind user’s ear by beaming live-streaming footage from the glasses' camera to the company’s agents who can then give audio instructions to the end users. (3)

Drawbacks:

The blind person should wait to be connected to the -Aira agents in order to be able to discover things around.

-Virtual Assistant as Support for People Visually Impaired

it aims, using artificial intelligence, the recognition of the denomination of paper currency, which could be complicated to do for people with disabilities since there is no Braille version of these; or the identification of faces, so when the blind person is in a room he or she can recognize if there are more people in there and, if the face of the person has been previously registered, the device will say with a voice who it is. (4)

Drawbacks:

-Not supporting Arabic Language

-due to the interaction that the user had to do with the touch screen of the mobile device to activate the application, focus the camera and take the photo, makes the process complicated.

- angles and lighting conditions affect the explicit recognition of the images.

Our project will improve the previous drawbacks, so we will focus on the blind and visually impaired together, so we want to add new features such as:

-sending a person's location to someone who wants the user to tell him where he is when he feels lost.

-Supporting Arabic reading and translation, not only English.

-The user can interact with the device with voice commands that is easier than using buttons.

-In keeping with the current time, we will also use a face shield as a body to protect the user from getting infected with Corona virus and become like any other person, as he does not wear anything strange in shape.

2. Analysis and Requirements

2.1. System Requirements

2.1.1. Enumerated Functional Requirements

	Requirements
REQ1	The text in the image is detected (using the camera) and converted to pure copy text (.txt).
REQ2	The text in the image is converted into speech/voice using the headphones.
REQ3	The text in the image is translated and converted into speech/voice using the headphones. the translation is done from Arabic to English and vice versa.
REQ4	The image of the person captured by the camera is recognized according to the saved database, and his name is output as a speech/voice
REQ5	The image of the currency captured by the camera is recognized according to the saved database, and its name is output as a speech/voice
REQ6	The image of the color of clothes captured by the camera is recognized, and color name is output as a speech/voice
REQ7	sending location of user to a specific person he chooses using the GPS module
REQ8	The user interacts with the device with his voice
REQ9	The user has the option to change the language from English to Arabic and vice vera

2.1.2. Enumerated Nonfunctional Requirements

	Requirements
REQ10	-Performance: the module must be able to perform properly and extract the text from any image, and also recognize correctly any person or currency or color, and also send the correct location of the user for specific person that the user chooses.
REQ11	- Reliability: the module must achieve high accuracy in identifying the text in the image and translating it and also high accuracy in identifying any person that stored in the database, and it should identify him if he isn't stored.
REQ12	-Flexibility: The module can be used by all blind people and people with vision problems and in different places such as college, school, hospital and even in the streets, it can be easy to be portable.
REQ13	-Usability: The raspberry pie module is light, safe, and easily wearable.

2.2. Functional Requirements Specification

2.2.1. Stakeholders

- End users (blind people or with vision difficulties).

- schools for blind.

- suppliers.

2.2.2. Actors

Actors: Any person or organization or external system that plays an important role in the system or interacts with the system.

The actors in our system are:

-**User** (blind person that uses the module).

-**Camera:** it takes an image, and then the system will perform a specific task according to the audio that is entered by the user.

-**Specific person:** This person will be chosen by the user to send his location for when he feels he is lost.

-GPS module:

This is hardware components that will be connected with the raspberry pi that enables the user to send his location for a person.

2.2.3. Use Cases

a. Use case Description

it contains the scenario(detailing) of each use case.

English speech:

When the user selects English language, It is built to treat with the user by voice messages in English language.

Arabic speech:

When the user selects Arabic language, It is built to treat with the user by voice messages in Arabic language.

Hear the detected text:

When the user says “reading” in English language or ”اقرا” in Arabic, the device takes a picture and converts the text in the picture into audio text.

Translate the text:

When the user says "Translate" in English language or ” ترجم” in Arabic, the device takes a picture and translates the text in the picture from English to Arabic or from Arabic to English and then converts the translated text into audio text.

Recognize person:

When the user says " recognition" in English language or “وجه” in Arabic, the device takes a picture of the person and recognizes him then out his name as a voice.

Recognize currency:

When the user says, "currency " in English language or” عملة” in Arabic, the device takes a picture of the currency and recognizes it then out its information and name as a voice

Recognize color:

When the user says, "Color" in English language or “لون” in Arabic, the device takes a picture of what he wants to know its color and recognizes its color then out the name of color as a voice

Send location:

When the user says "location" in English language or “موقع” in Arabic, the device sends his location to a specific person.

Language:

When the user says” language "in English language or” لغه” in Arabic language, he has the option to select English language or Arabic language to treat with the module and there is default option (Arabic language) if he didn’t the language and this is the first time for him to use the module

b. Use Case Diagram

The use case diagram is a behavior diagram that describes the functional requirements of the system, it shows a set of use cases, and how the actors use them.

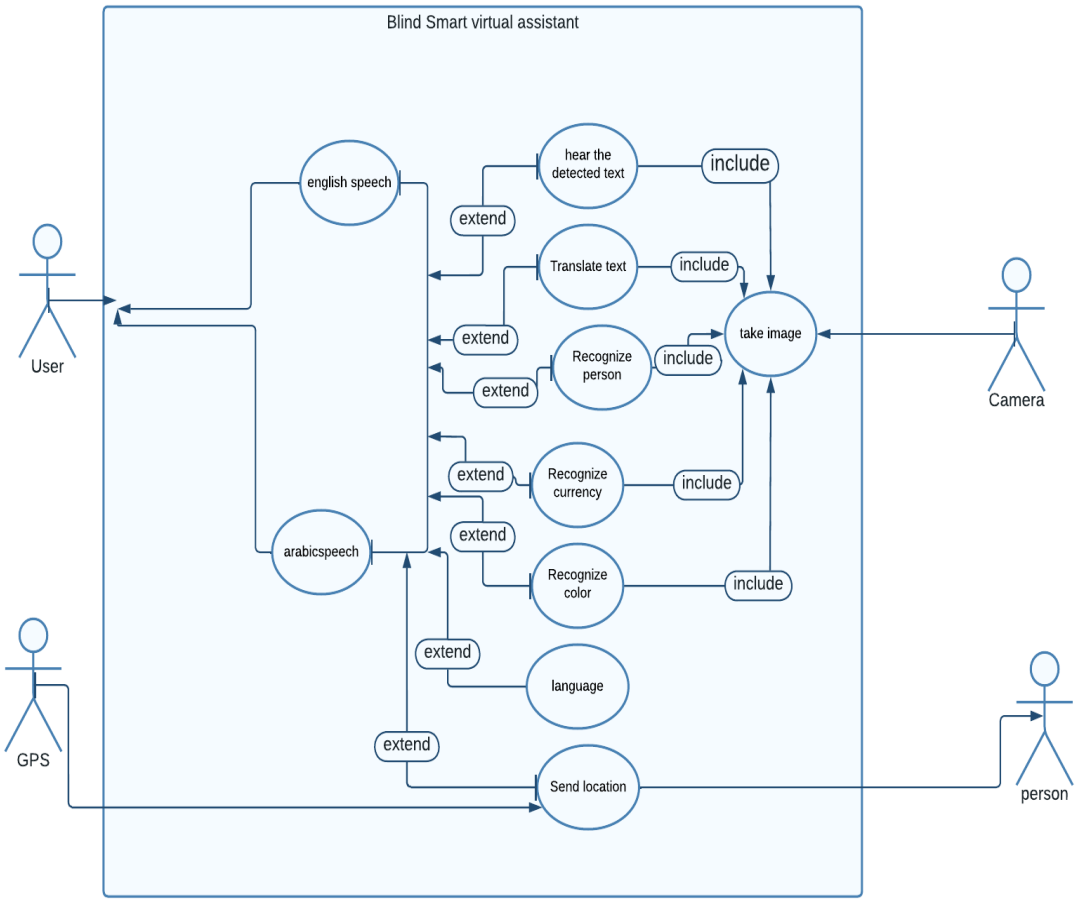

Figure4 use case diagram

2.2.4. System Sequence

Sequence diagram is a dynamic model of the use case, showing the interaction among classes during a specified time period.

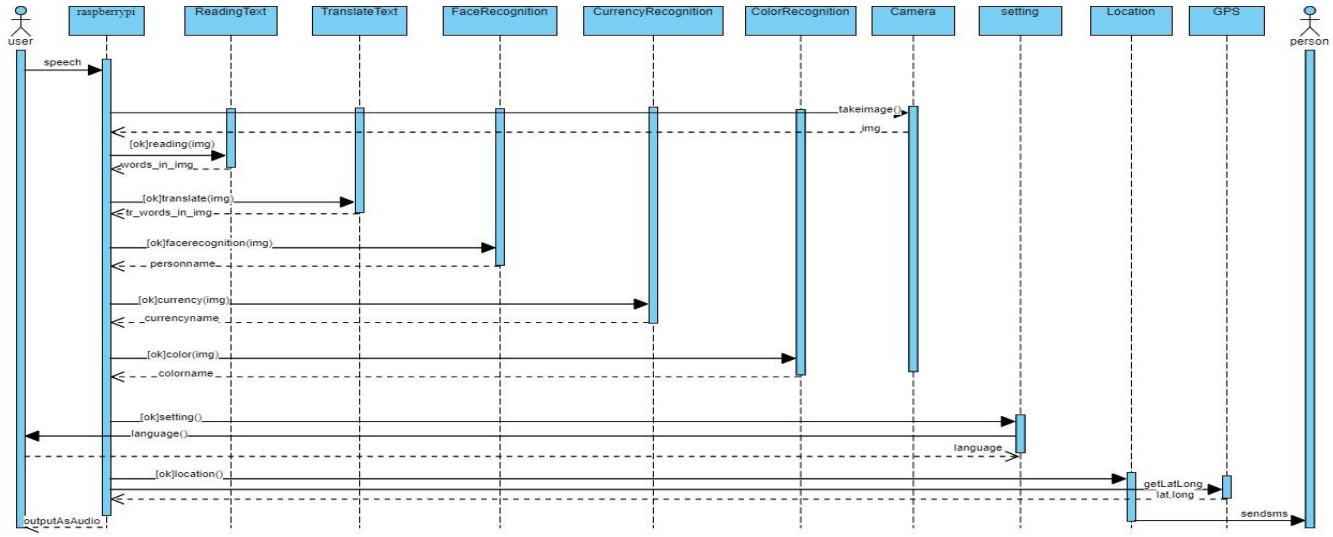

Figure5 sequence diagram

2.2.5. Class Diagram

Describes the static structure of the system more than the behavior

Operation: is the implementation of any service that can be requested from the system, each method(operation) describes what the class can do.

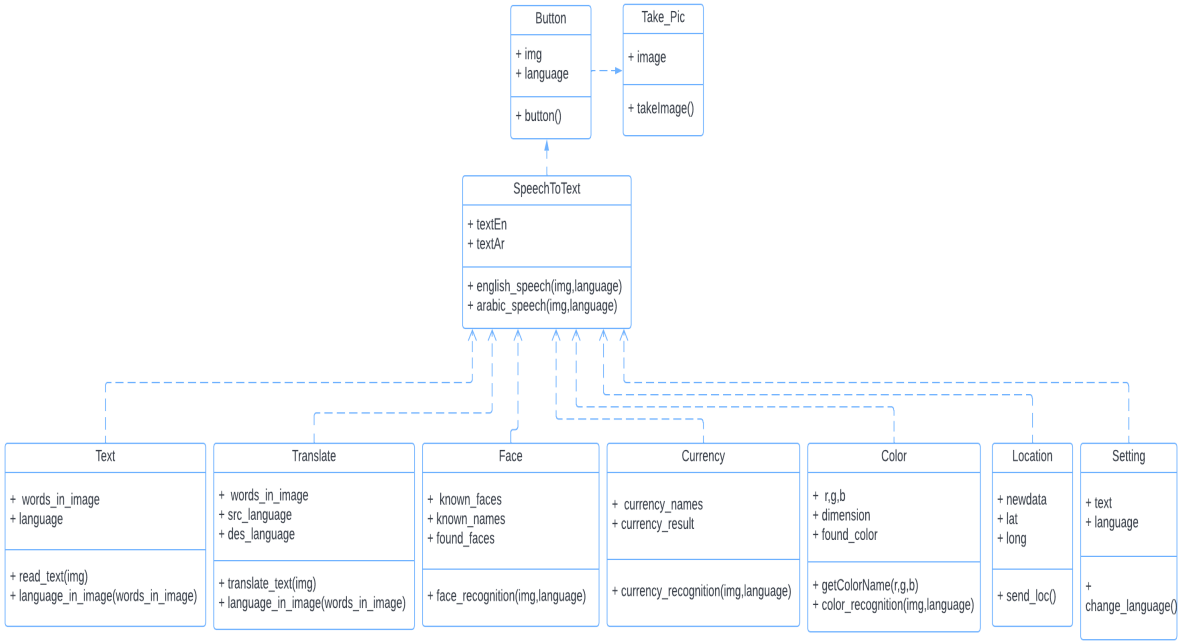

Figure6 Class diagram

2.2.6. Block diagram

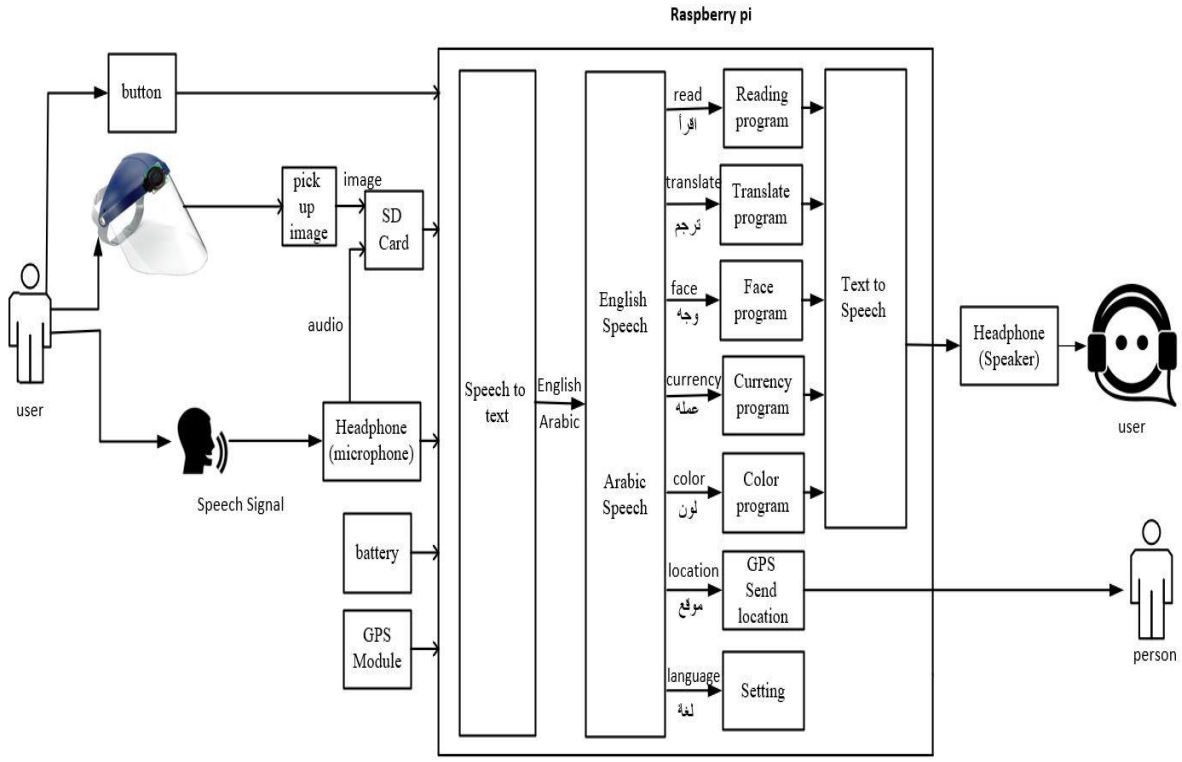

Figure 7 Block diagram

First , when the user wants the module to perform a specific task , he will push the button , so the camera(web cam with resolution 1080 pixels) that contained in the module will pick up an image that will be stored in the SD card which is a storage for the raspberry pi, and then the module will out a voice message that requires the user to enter what he wants(which feature) "what do you want " if the language is English or " كيف يمكنني ان اساعدك" if the language is Arabic (the user has two options English language or Arabic language) , the module will treat with the user with this language to perform the tasks , so if the language is English language the module will require from the user to select task to be performed (language, Reading , Translate , Recognition , currency, color, location) and the voice message the user entered will be processed(converted into text) using speech recognition technique and then the raspberry pi will perform the function according to the voice message of the user and outs voice message of the result for him , and the same scenario will be performed if the user selects Arabic language but the treating with the user will be performed in Arabic language.

The module treats the user according to the current language that is stored, if it is English so the module treats the user in English language and if it Arabic the module treats the user in Arabic language ,and the user has the option to change it , and if the user has the module and this is for the first time for him to use it and he didn't select the feature "language " , so the default option is to be Arabic.

So the sequence of the module is that, the module will be autorun as soon as it connects to power, so the user need to press the button so the audio message to the user that the module is ready to take an image, and then, the camera will capture an image , and according to the current language that is in the lang.txt file, the module will require from the user to select the task , he wants the module to perform, after any success trial , the module outs an audio message that tells the user he could use it another time to perform any task he wants from the module (" the process finished") in English or " يمكنك المحاوله مره اخري " in Arabic.

We have a module (Raspberry pi) that helps the person to listen to a voice that helps him to hear the detected converted text or the words that translated, recognize people, recognize the currency

3. Testing and Validation

3.1. Test Cases

valid/ invalid	Test data	Expected result	Actual result	pass/ fail
Valid Data	Saying “Reading” Image of text	The audio of the text	The audio of the text	Pass
	Saying “اقرا” Image of text	The audio of the text	The audio of the text	Pass
	Saying “Translate” Image of text	The audio of the translated text	The audio of the translated text	Pass
	Saying"ترجم" Image of text	The audio of the translated text	The audio of the translated text	Pass
	Saying “recognition” Image of person	name of person in English	name and of person in English	Pass
	Saying”وجه “ Image of person	name person in Arabic	name of person in Arabic	Pass
	Saying” currency “ Image of currency	name of currency in English	name of currency in English	Pass
	Saying"عمله" Image of currency	name of currency in Arabic	name of currency in Arabic	Pass
	Saying” colors” Image of clothes	name of color in English	name of color in English	Pass
	Saying"لون" Image of a person	name of color in Arabic	name of color in Arabic	Pass
	Saying "location	Location of the GPS module	Location of the GPS module	Pass
Invalid Data	Saying “Reading” image without text	Failed detect	Failed detect	Pass
	Saying “Translator” Image of text	Failed detect	Failed detect	Pass
	Saying”وجه “ Image of text	Failed detect	Failed detect	Pass
	Saying “currency “ Image of color	Failed detect	Failed detect	Pass
	Saying"عمله" Image of a person	Failed detect	Failed detect	Pass
	Saying “face “	Voice message says” unknown person “.	Voice message says” unknown person “.	Pass

	image of person but not stored in the database.			
	Saying "Reading", text image but the language of image isn't English nor Arabic	Failed detect	Failed detect	Pass
	Saying "color", image of a person	Failed detect	Failed detect	Pass
	Saying "object", image of currency	Say "invalid word", please try again"	Say "invalid word", please try again	Pass

3.2. Unit testing

Setting unit

Test case #	Test case description	Test data	Expected result	Actual result	Pass/ fail
1	Check if the user says any other language except English or Arabic	“Spanish” language, Text image	Voice message that tells the user that this is an invalid word.	Voice message that tells the user that this is an invalid word.	pass
2	Check if the user says correct language "English "or Arabic"	“English” language	Voice message that asks the user to if he wants to speak English or Arabic	Voice message that requires the user to say the language he wants (English or Arabic)	pass
3	Check if the user says correct language "English "or Arabic	“Arabic” language	Voice message that asks the user to if he wants to speak English or Arabic	Voice message that requires the user to say the language he wants (English or Arabic)	pass

Text unit

Test case#	Test case description	Test data	Expected result	Actual result	Pass/ fail
1	Checks if the image that entered to the system valid (Arabic language) and valid language, and valid command	Arabic Text image, “Arabic language," اقرا "command	Audio message that contains the Arabic text that contained in image	Audio message that contains the Arabic text that contained in image	pass
2	Checks if the image that entered to the system valid English language) and valid language, and valid command	English Text image, "English language, "Reading "command	Audio message that contains the English text that contained in image	Audio message that contains the English text that contained in image	pass
3	Check if the image that is entered to the module doesn't contain text	Image of person image, English language "Reading "command	Failed to detect	Failed to detect	pass

Translate unit

Test case#	Test case description	Test data	Expected result	Actual result	Pass/ fail
1	Checks if the image that entered to the system valid (Arabic language) and valid language, and valid command	Arabic Text image, “Arabic language," ترجم "Command	Audio message that contains the translated text in English language that	Audio message that contains the translated text in English language that	pass

2	Checks if the image that entered to the system valid English language) and valid language, and valid command	English Text image, "English language, "Reading "command	Audio message that contains the translated text in Arabic language	Audio message that contains the translated text in Arabic language	pass
3	Check if the image that is entered to the module doesn't contain text	Image of currency image, English language "Reading "command	Failed to translate	Failed to translate	pass

Face unit

Test case#	Test case description	Test data	Expected result	Actual result	Pass/ fail
1	Check if the user enters valid data	Image of a person stored in database, English language, Face command	Audio message that contains the name of the person that the user wants to recognize	Audio message that contains the name of the person that the user wants to recognize	Pass
2	Check if the user enters valid data	Image of a person English language, Face command but this person isn’t stored in the database	Failed to recognize	Failed to recognize	Pass
2	Check if the user enters invalid data	Image of color English language, Face command	Failed to recognize	Failed to recognize	Pass

Currency unit

Test case#	Test case description	Test data	Expected result	Actual result	Pass/ fail
1	Check if the user enters valid data	Image of a currency stored in database, English language, currency command	Audio message that contains the amount of the currency that the user wants to recognize	Audio message that contains the amount of the currency that the user wants to recognize	Pass
2	Check if the user enters valid data	Image of a person English language, currency command but this currency isn't stored in the database	Failed to recognize currency	Failed to recognize currency	Pass
3	Check if the user enters invalid data	Image of color English language, currency command	Failed to recognize	Failed to recognize	Pass

Color unit

Test case #	Test case description	Test data	Expected result	Actual result	pass / fail
1	Check if the user enters valid data	Image of color, English language, color command	Audio message in English that contains the color or colors that contained in the image	Audio message that contains the color in the image the user wants to recognize	pass
2	Checks if the user enters valid data	Image of color, Arabic language, " "لون command	Audio message in Arabic that contains the color or colors that contained in the image	Audio message that contains the color in the image the user wants to recognize	pass
3	Checks if the user enters invalid data	Image of color, Arabic language, "color" command	Failed to detect	Failed to detect	pass

4. Tools and Technologies

4.1. Technologies

Deep learning: Using techniques of deep learning for creating models

Python: The programming language used

The software resources that We used python programming language to program these features and its tools (libraries) as:

- •Pytesseract for text detection(OCR)
- •gTTS for (text to audio)
- •google trans (for text translation)
- •Face recognition, dlip
- •ORB (for Currency detection and recognition)
- •Matplot, pandas, numby for color recognition
- •speech recognition for speech to text

and the IDE is Visual Studio Code.

4.2. Tools

4.2.1. Software:

a) Speech recognition

The **first** component is speech. It must be converted from a sound to a signal that can travel through a microphone and can be transcribed to digital data. This is done using an analog to digital converter. Once the form of data is digitized, several trained models can easily transcribe the audio to text.

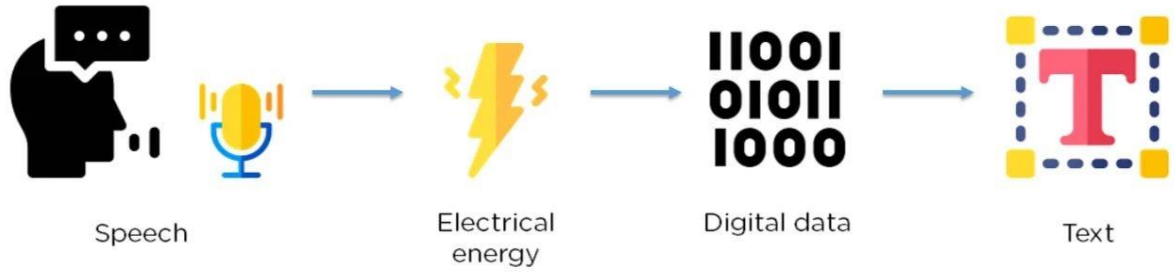

Figure 8 first phase in speech Recognition

then Apply the Automatic Speech Recognition

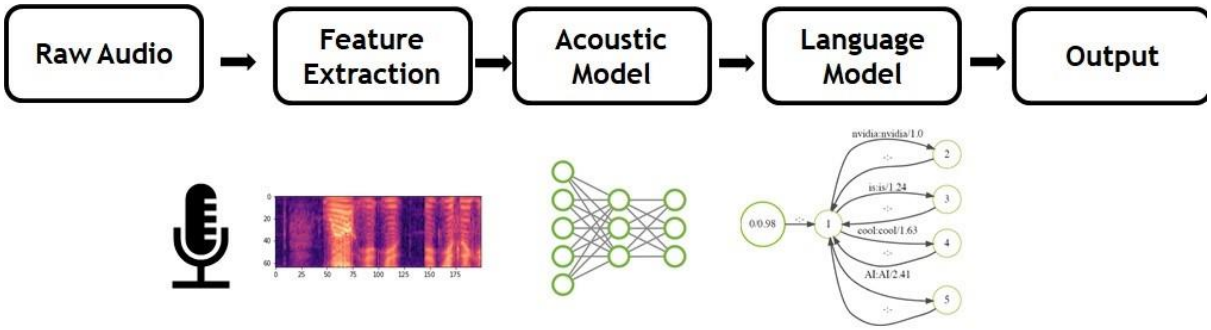

Figure 9 second phase in speech recognition

As we are aware sound is nothing more than vibrations of the air that humans are trained extremely well to decode, we have a stream of words that a person has uttered

The goal of the step is to have chunks of speech, we can also hear is phoneme

Acoustic model is the part of ASR responsible for mapping sound to phones is called

The acoustic model as a set of building blocks, boxes which contain models for all phones in each language, there are boxes labelled A, B, C and so on, depending on which phones are used in the particular language

On top of that, part of this construction set are also contextual probabilities this means how likely a phone is to follow another

This is rendition of the sound wave that my mouth produced when I spoke. The task of the acoustic model is to guess which phones I pronounced and how do they combine to a word

The acoustic model processes the sound and compares it to the models of individual phones from its boxes, The chunks that I uttered will be like more than one box

The acoustic model takes this into account and looks at the neighboring chunks and their contextual probabilities, outputs the most likely result, a stream of phones

the language model for mapping phones to words and phrases

The language model as a huge table that contains the probabilities of two words following each other

b) OCR (optical character recognition)

Before all, let me define what machine learning is:

Machine learning is a field of study that gives computers the ability to learn without being explicitly programmed.

The main purpose of OCR is to detect the text in an image and convert it to transcription text or (.txt) document.

So, this process consists of subprocesses or modules, each module acts a task on some piece of data and delivers to the next module to perform its task and so on(pipeline), and each module of these is a machine learning component (5)

-The Process :

image ⟶ text detection ⟶ character segmentation ⟶

character recognition

These modules are:

1. Text detection: the purpose of this module (machine learning component) is to determine the regions of image where there is text.

Figure 10 image with text

This done using supervised learning algorithm.

The difference between Supervised learning algorithm and unsupervised learning algorithm:

The main distinction between the two approaches is the use of labeled datasets. To put it simply, supervised learning uses labeled input and output data, while an unsupervised learning algorithm does not. While supervised learning models tend to be more accurate than unsupervised learning models,

the training of the machine is done that:

Patches of images that contain text are positive examples(y=1).

Patches of images that doesn't contain text are negative examples(y=0).

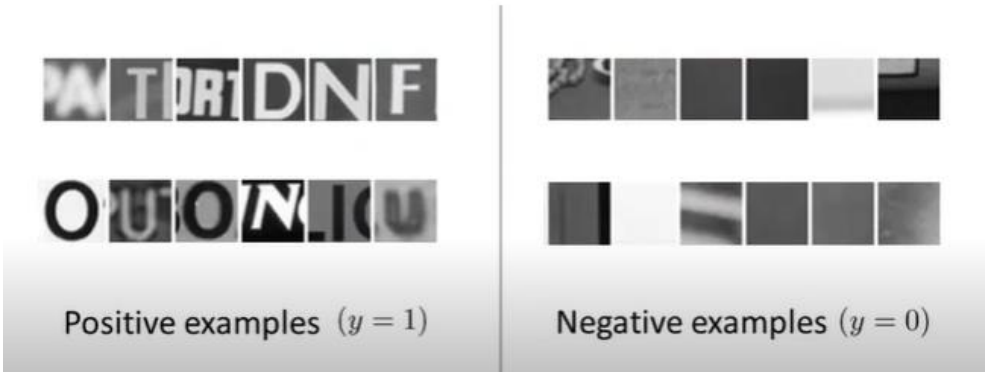

Figure 11 positive and negative examples

When testing the machine, it can detect the regions of image that contain text.

2. Character segmentation:

The task or purpose of this module (machine learning component) is to separate out characters. This is done using that: There is one dimensional sliding window from left to right, when there

Figure 12 character segmentation

is gab or midpoint, so these are two distinct characters, and they are separated, if else, so it is one character and isn't separated.

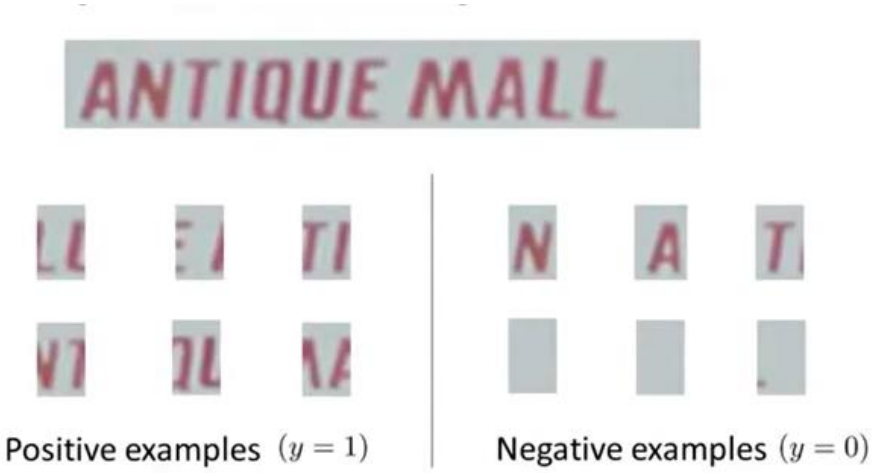

Figure 13 positive and negative examples

3. Character recognition/classification:

Figure 14 character classification

The purpose of this module is to recognize or classify each individual character and recognize which alphabet from 26 in English language and from 28 in Arabic language.

c) GTTS (google text to speech)

Text to speech is a mapping problem from sequence to sequence (from word / text to audio signal / acoustic waveforms).

The text is to be synthesized and converted into speech waveforms.

This task is done using the neural network.

How humans can talk!

The text or concept in our brain is to be synthesized and converted into the movement of our muscles, and using the air from lungs, we create vocal source excitation signal) using the vocal cord. then there is frequency characteristics by the vocal transfer function controlled by articulators, then we emit the speech from our mouse.

so the aim of text to speech synthesis is to mimic this process by the computer in some way.

Google text to speech (gtts)

In needs an active internet connection because it uses google API for text to speech translation.

- text to speech translation is performed in two phases:

i. natural language processing (NLP)

This phase for pre-processing of the text that is passed :(front end).

ii. Digital signal processing (DSP).

And this for convert this text to speech :(back end)

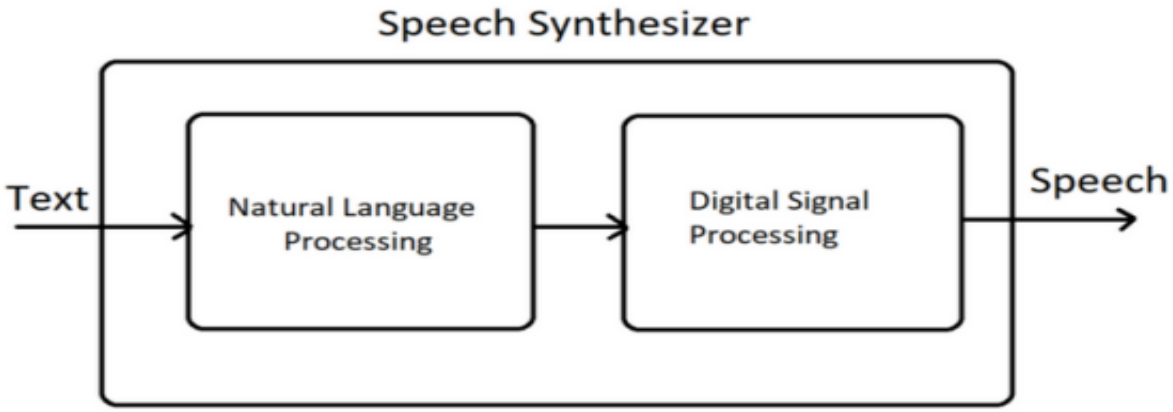

Figure 15 speech synthesizer

A text-to-speech system is composed of two parts: a front-end and a back-end.

-The front-end has two major tasks:

First, it converts raw text containing symbols like numbers and abbreviations into the equivalent of written-out words. This process is often called text normalization, pre-processing, or tokenization. The front-end then assigns phonetic transcriptions to each word, and divides and marks the text into prosodic units, like phrases, clauses, and sentences. The process of assigning phonetic transcriptions to words is called *text-to-phoneme* or *grapheme-to-phoneme* conversion. Phonetic transcriptions and prosody information together make up the symbolic linguistic representation that is output by the front-end.

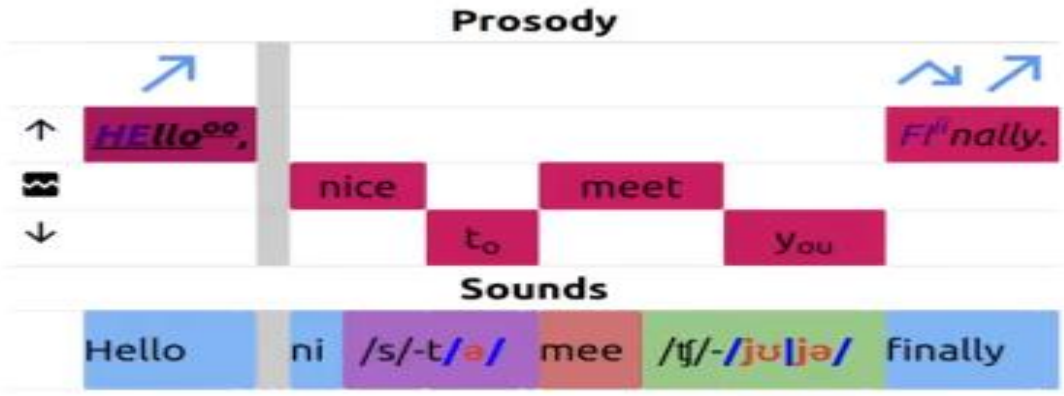

Figure 16 prosody

-The back-end—often referred to as the *synthesizer*—then converts the symbolic linguistic representation into sound. In certain systems, this part includes the computation of the *target prosody* (pitch contour, phoneme durations), which is then imposed on the output speech.

d) google trans API

- word-word translation:

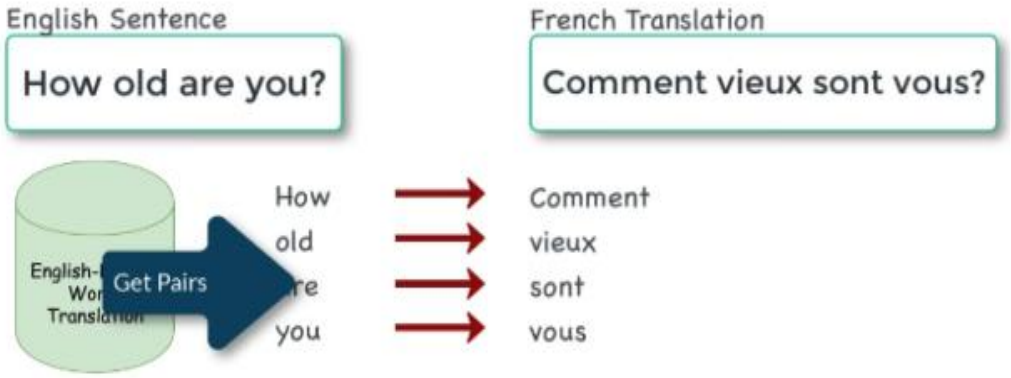

Figure 17 word-word translation

we have a sentence in one language and wants to translate this

sentence to another language, this is done as this sentence is segmented into words and each word is translated from the source language to the corresponding target language using

the database of each language:

∵ it is a beautiful day.

If this is the sentence and we want to translate it into Arabic, this segmented out as:

هذا It

يكون Is

A

جميل beautiful

يوم Day

But in any language, there are two important components:

1- tokens:

Smallest units of the language.

2- Grammar:

Defines how these tokens should appear. (it is a guide or set of rules that defines the ordering of these words).

If the translation matters only about the tokens and doesn't matter about the grammar, the translation would be so simple, but isn't the case, the grammar should be considered unless this translation will produce gibberish language. So, what is the grammar cares about

1-syntax:

checks if the structure of the translated sentence is correct or not.

2-sematics analysis:

Checks if the sentence makes sense in context. To understand the grammar, we will let the neural network do it. Neural network is a component that learns how to solve a problem by looking at hundreds of thousands of examples. We will show the architecture of neural network based on the problem we are trying to solve. Any sentence consists of sequence of words, the human understands that but neural network doesn't understand that so we should convert it into a form that neural network understands.

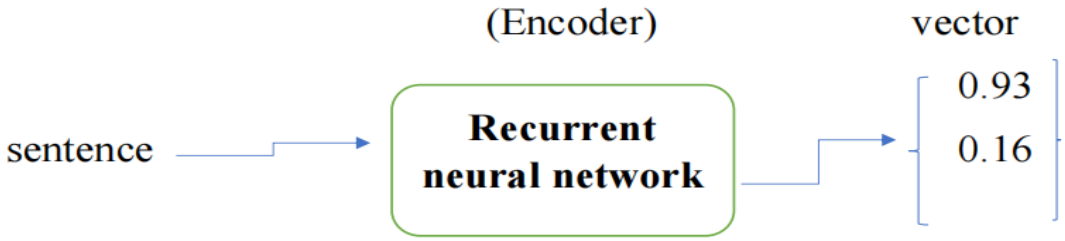

Figure 18 Encoder

recurrent neural network:

A recurrent neural network (RNN) is a type of artificial neural network which uses sequential data or time series data.

recurrent neural networks utilize training data to learn. They are distinguished by their "memory" as they take information from prior inputs to influence the current input and output.

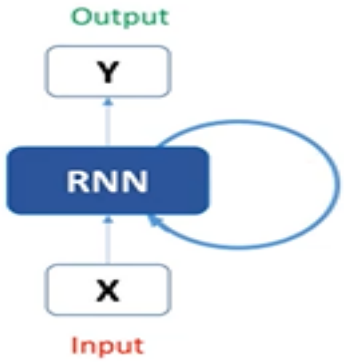

Figure 19: RNN

These algorithms are commonly used for ordinal or temporal problems, such as language translation, natural language processing (nlp), speech recognition

RNN takes a sentence and converts it into a vector of numbers.

Recurrent Neural Network vs. Feedforward Neural Network

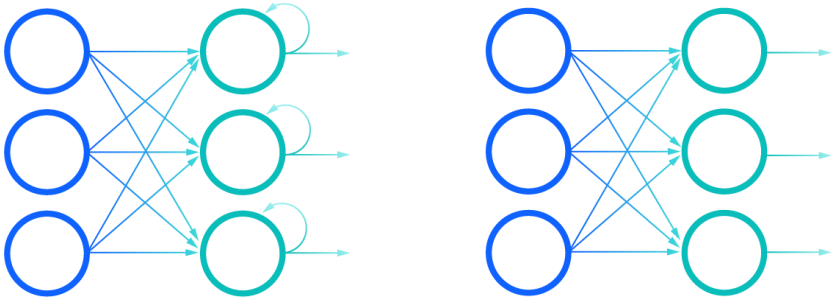

Figure 20: forward neural network vs recurrent neural network

- We want to convert this vector (computer data) into the target language, so it is done using another recurrent neural network.

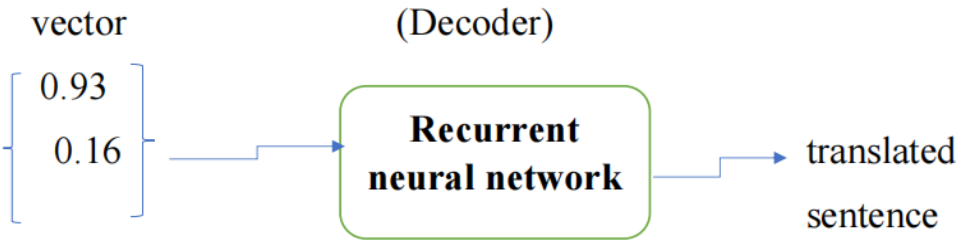

Figure 19 Decoder

so 2 recurrent neural networks are the architecture of language translator,

It is also called Encoder-Decoder architecture.

The first recurrent neural network is Encoder: Because it encodes the sentence to the computer data.

The second recurrent neural network is Decoder:

Because it decodes the computer data into the translated sentence.

Recurrent neural network (LSTM -RNN):

Long short-term memory recurrent neural network, this is a popular RNN architecture, which was introduced by Sepp Hochreiter and Juergen Schmidhuber as a solution to vanishing gradient problem, they work to address the problem of long-term dependencies. That is, if the previous state that is influencing the current prediction is not in the recent past, the RNN model may not be able to accurately predict the current state, LSTMs have "cells" in the hidden layers of the neural network, which have three gates–an input gate, an output gate, and a forget gate. These gates control the flow of information which is needed to predict the output in the network.

But there was a problem:

When the translator translates a word to the target language, it looks at only the previous word in the source language,

Recurrent neural network

although, the word depends on the previous word and the next word, so the recurrent neural network was replaced with

bidirectional neural network. And to improve the translation," the attention mechanism" was added between the encoder and the decoder (to make the alignment between the I/p and the o/p). And after this to improve the translation, the architecture of the decoder and the encoder became 8LSTM to understand the grammar and semantics of the language. So finally, the architecture of the translator is:

1-Encoder (8 LSTM -RNN)

2- attention mechanism

3-Decoder (8 LSTM -RNN)

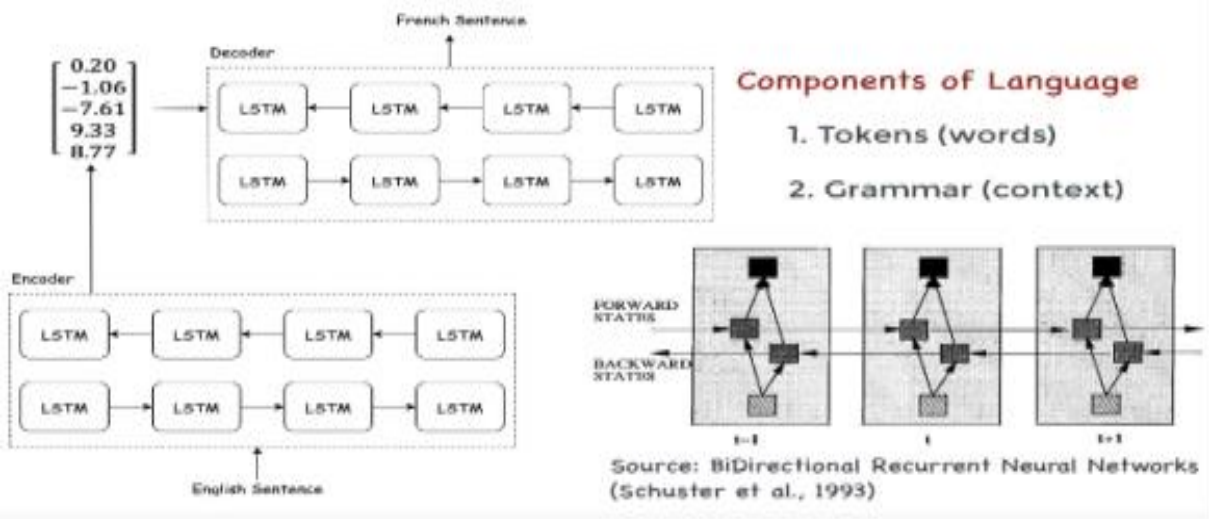

Figure 20 translator architecture

The operation of translation is done as that:

The source sentence is passed to the encoder that converts each word in the sentence into the corresponding computer data, (each word has a vector of computer data), and then these computer data are passed to the attention mechanism to determine which word the source sentence should be focused on while generating the translated sentence, and then the decoder will take these words (from the attention mechanism) to consist the translated sentence. This happens each time we use google trans(6)

e) Face recognition

1- detect the face in the photo:

We've used Dlib's HOG (Histogram of Oriented Gradients) and Linear SVM as face detection algorithm.

Dlib was originally introduced as a C++ library for machine learning by Davis King. But later, a Python API was also introduced which we can easily install using pip install dlib

Dlib provides the get_frontal_face_detector() function which instantiates the face detector for us ,

next steps will show how it work:

Step 1: Preprocessing:

Figure21 resize image

We need to preprocess the image and bring down the width to height ratio to 1:2. The image size should preferably be 64 x 128. This is because we will be dividing the image into 8*8 patches to extract the features.

Step 2: Calculate the Gradient Images

Gradients are the small change in the x and y directions. so, it takes a small patch from the image and calculate the gradients

it selected pixel, then to determine the gradient in the x-direction, we need to subtract the value on the left from the pixel value on the right. Similarly, to calculate the gradient in the y-direction, we will subtract the pixel value below from the pixel value above the selected pixel.

This process will give us two new matrices – one storing gradients in the x-direction and the other storing gradients in the y direction. This is like using a Sobel Kernel of size 1. The magnitude would be higher when there is a sharp change in intensity, such as around the edges.

The same process is repeated for all the pixels in the image.

we can find the magnitude and direction of gradient using the Equation

$$g = \sqrt{g_x^2 + g_y^2}$$
$$\theta = \arctan \frac{g_y}{g_x}$$

Figure 22

Step 3: Calculate Histogram of Gradients in 8×8 cell

the image is divided into 8×8 cells, and the histogram of oriented gradients is computed for each cell.

By doing so, we get the features (or histogram) for the smaller patches which in turn represent the whole image.

If we divide the image into 8×8 cells and generate the histograms, we will get a 9 x 1 matrix for each cell.

Step 4: 16×16 Block Normalization

Although we already have the HOG features created for the 8×8 cells of the image, the gradients of the image are sensitive to the overall lighting. This means that for a particular picture, some portion of the image would be very bright as compared to the other portions

so, this lighting variation reduced by normalizing the gradients by taking 16×16 blocks.

by combining four 8×8 cells to create a 16×16 block. And each 8×8 cell has a 9×1 matrix for a histogram. So, we would have four 9×1 matrices or a single 36×1 matrix. To normalize this matrix,

The resultant would be a normalized vector of size 36×1. (7)

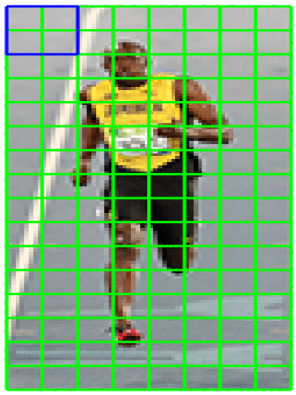

Figure 23 Block normalization

2- Locate the 68 Facial Landmarks

it uses shape_predictor_68_face_landmarks.dat it is pre-trained facial landmark detector that detects 68 different facial landmarks

Figure24 68_facial Landmarks

3- Create a 128-value encoding of the facial landmarks

It uses dlib_face_recognition_resnet_model_v1.dat this generation of the 128 encoding values is done with a pre-trained CNN, the network architecture for face recognition is based on ResNet-34 from the Deep Residual Learning for Image Recognition paper by He et al., but with fewer layers and the number of filters reduced by half.

The network itself was trained by Davis King on a dataset of ≈3 million images.

4- comparing faces

to know how similar the faces are, so we're going to compute the Euclidean distance np.linalg.norm(face_encodings - face_to_compare, axis=1)

it takes each point, 1 through 128, and you subtract the second one from the

first one, you square each one, you sum them together, and take the square root, and that's going to give you a single number. That number is going to be used to determine whether these two images are of the same person's face. if it is <0.6, the faces match, and if it is >0.6, the faces are different

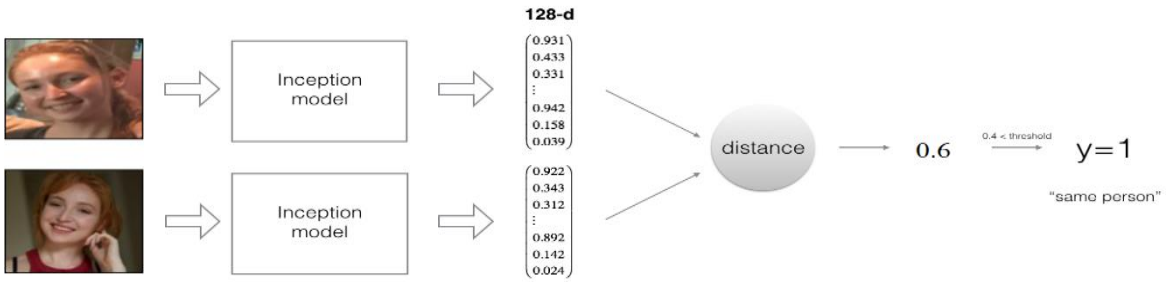

Figure25 comparing faces

f) ORB (for Currency detection and recognition)

ORB algorithm is proposed based on FAST algorithm and BRIEF algorithm.

1- Fast

FAST The Features from Accelerated Segment Test (FAST) algorithm works in a clever way.

it draws a circle around including 16 pixels. It then marks each pixel brighter or darker than a particular threshold compared to the center of the circle. A corner is defined by identifying several contiguous pixels marked as brighter or darker.

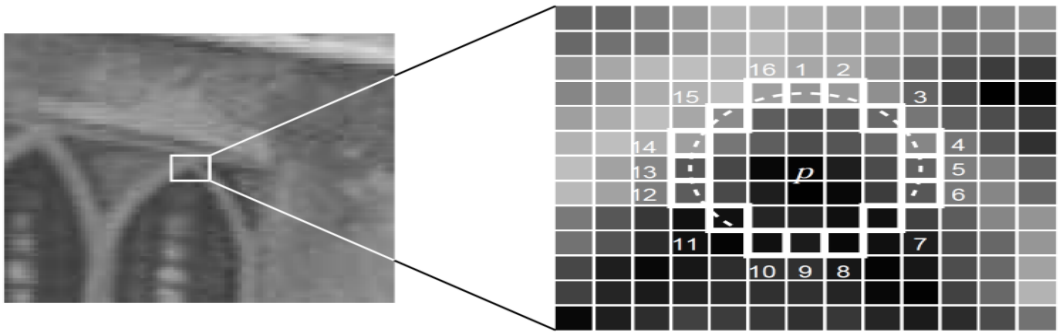

Figure 26 extract important features

FAST implements a high-speed test, which attempts at quickly skipping the whole 16-pixel test.
three out of four of the test pixels (pixels number 1, 9, 5, and 13) must be within (or beyond) the threshold (and, therefore, marked as brighter or darker) and one must be on the opposite side of the threshold. If all four are marked as brighter or darker, or two are and two are not, the pixel is not a candidate corner.

FAST is an incredibly clever algorithm, but not devoid of weaknesses, and to compensate these weaknesses, developers analyzing images can implement a machine learning approach, feeding a set of images (relevant to your application) to the algorithm so that corner detection is optimized. (8)

2- BRIEF

BRIEF extracts descriptors around feature points by binary coding method.

Around the image spot P is, randomly selecting n pairs of pixel point and defining it as

$$\tau(p:x,y)=\begin{cases} 1 & \text{if } p(x)<p(y) \\ 0 & \text{otherwise.} \end{cases}$$

g) Color recognition

The panda's library(9) is very useful when we need to perform various operations on data files like CSV. pd.read_csv() reads the CSV file and loads it into the pandas DataFrame. We have assigned each column with a name for easy accessing.

We have the r,g and b values of colors in CSV file, and we created a set for dimensions of x and y ,to compare its r ,g, b with r,g,b in csv file to get the color of this pixels

To get the color , we calculate a distance(d) which tells us how close we are to color and choose the one having minimum distance.

Our distance is calculated by this formula:

d = abs(Red – ithRedColor) + (Green – ithGreenColor) + (Blue – ithBlueColor)

4.2.2. Hardware:

The idea of the project is to make a module for blind people that help them in education life at university, school and make life easier while treating with people to recognize them and doesn't need assistance all time from the people around him. So, based on the goal of the project and after doing some searches, team members decided to work with the following components:

a) Raspberry Pi3 Model B+ (Controller)

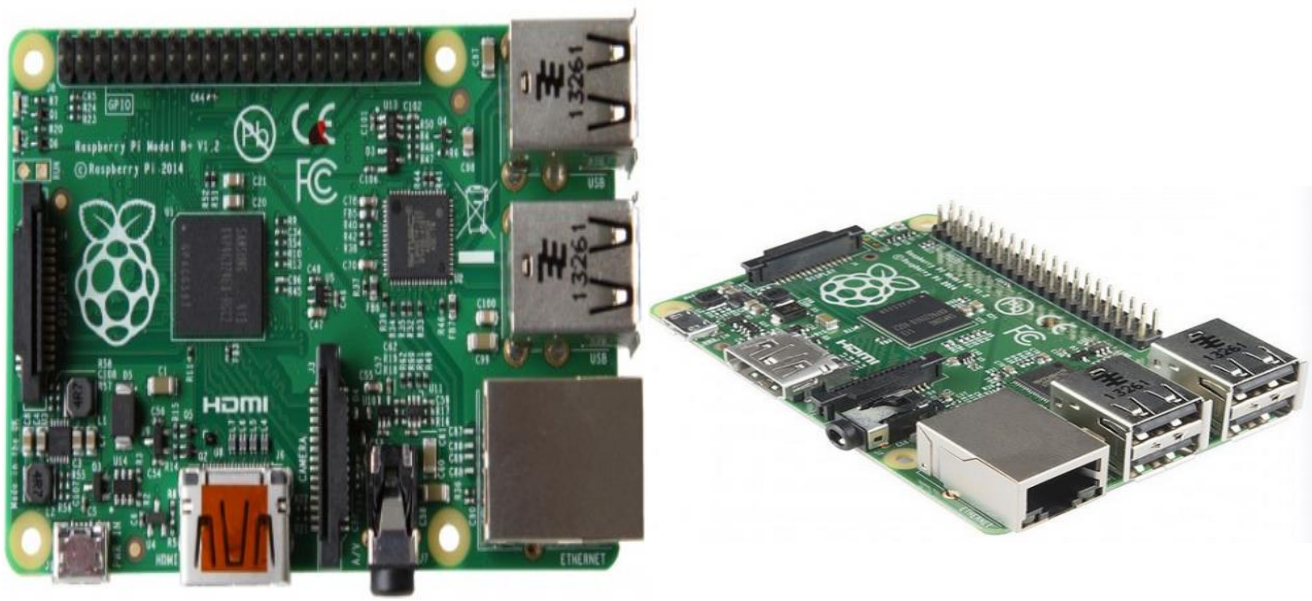

Figure27 Raspberry Pi3 Model B+

The Raspberry Pi is a credit card-sized computer with an ARM processor that can run Linux.

we needed a few additional things that are not included:

-A 5 V power source with a micro-USB connector (we take the power of raspberry pi from laptop)

- A microSD card with an operating system on it, which serves as the main storage of the raspberry pi.

Input and output devices, such as a keyboard and display(monitor) with a HDMI cable instead we enabled the VNC to access the raspberry from our laptop,

raspberry pi also includes GPOI (general purpose input/output) pins to control electronics components.

We also considered getting a case to house and protect our Raspberry Pi.

Figure28 Case of raspberry pi

Function of raspberry pi:

Raspberry Pi 3 used for many purposes such as education, coding, and building hardware projects. It is the main component of the project. It used as a low-cost embedded system to control and connect all of the components together, It uses the Raspbian, Linux or ubuntu as the operating system which can accomplish many important tasks However, for the project we decided to work on Linux as the operating system for raspberry pi. (10)

b) GPS NEO-7M Module

The Global Positioning System (GPS) is a satellite-based navigation system made up of at least 24 satellites. GPS works in any weather conditions, anywhere in the world, 24 hours a day, with no subscription fees or setup charges.

GPS satellites circle the Earth twice a day in a precise orbit. Each satellite transmits a unique signal and orbital parameters that allow GPS devices to decode and compute the precise location of the satellite. GPS receivers use this information and trilateration to calculate a user's exact location. Essentially, the GPS receiver measures the distance to each satellite by the amount of time it takes to receive a transmitted signal. With distance measurements from a few more satellites, the receiver can determine a user's position.

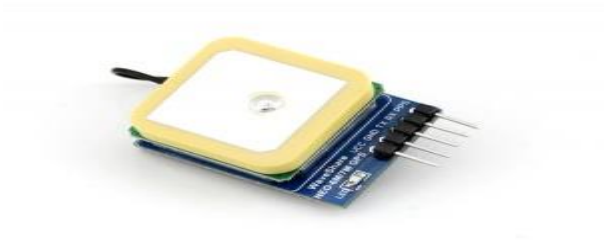

Figure29 GPS NEO-7M Module

-Vcc of GPS module Connected to Power Supply Pin 3 (5V) of Raspberry Pi.

-Tx (Transmitter Pin) of GPS module Connected to Pin 6 of Raspberry Pi.

-GND (Ground Pin) of GPS module Connected to Pin 5 Raspberry Pi.

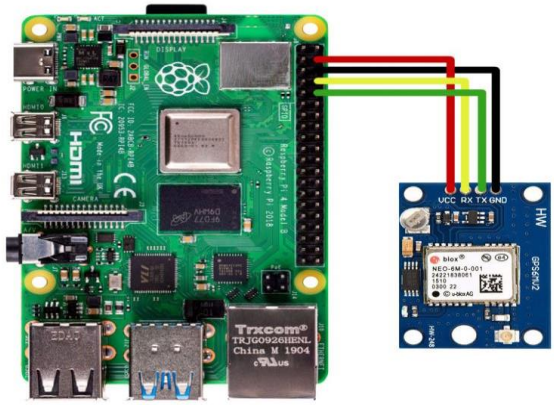

Figure30 connecting GPS module with raspberry pi

c) Web cam

it can capture images with maximum resolutions of 1080 pixels. It is compatible with most operating systems platforms such as Linux, Windows, and Mac OS. It has a USB port.

Figure31 webcam

Its function in the project:

In the project, the Webcam will be used the eyes of the person who wears the "Blind Smart virtual assistant." The camera is going to capture a picture when the button is pressed, in order to detect and recognize the text from the image

d) Pushbutton

Figure32 Bush button

e) SD Cards 32GB

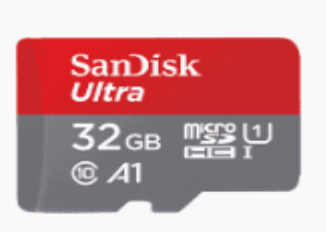

Figure 33 SD card

Raspberry Pi 3 uses a micro-SD card for storage (OS, libraries, and user programs).

f) USB sound adapter

A USB audio adapter is a cheap yet pragmatic piece of hardware to enhance the audio quality output of the laptop or desktop computers. Truthfully, the USB audio adapters can bring out the best in the headphones.

In our project, we used it to input voice and output voice to and from the raspberry pi at the same time

Figure 34 USB Sound adapter

g) headphones

Wired headphones are used in the project since Raspberry Pi 3 Model B+ come with Audio jack, it is better to take advantage of this feature

Figure35 headphones

Function of headphones:

The headphones will be used to help the user listen to the text that is been converted to audio after it has been captured by the camera or to listen to the translation of the text or listen the name of person he wants to recognize or color or currency, The headphones are small, light, and they are connected to the raspberry by its audio jack.

h) Face shield

Figure36 Face shield

The purpose of using the face shield is to protect the user from viruses especially with the spread of viruses such as corona virus, so it makes the module more protective

i) jumpers

Figure37 jumpers

For connecting the button with the raspberry pi.

5. Module Design and Implementation

At the early stage, the initial idea of the project was to create “smart module for blind” that capture any text images and convert it to voice, then, if the user wants to translate the text, he/she can say “translate” to translate any word he cannot understand from English to Arabic or vice vera. Initially, the team chose Linux as an operating system to be installed on the Raspberry Pi B+ and implement all the module functions. Also, the camera was built in the module to capture the images. The main target for the “Blind Smart virtual assistant” was blind people to allow them conceptually to use it anytime anywhere such as:

- Schools
- Universities
- Hospitals
- institutes
- Road/ Streets

To accomplish the project, at the conceptual design the team proposed the initial hardware below:

- Raspberry pi3 Model B+
- Camera
- headphones
- buttons
- SD card

After that, we wanted to create a module that solves many previous problems and at the same time be safe and protective for the user and support many features for him to make his life as easy as possible and easy as normal people.

So, from here we decided that our module will support:

-Arabic language aside from English language

- the interaction between user and the module will be using voice commands

- to guarantee the safety of the user, we decided to use GPS module and connect it with the raspberry pi so that the user can send his location to someone when he feels he is not safe or lost.

-also, with the spread of viruses, we suggested that the outside look module will be face shield to ensure the user will be protected.

However, during the work on module, some improvements and modifications are introduced such as the Raspberry Pi has been changed from putting it on the face shield to put it outside because it is large, and it is not comfortable for the user to wear it on the user's head. So, the user will wear the Raspberry Pi 3 Model B+ on his/her upper arm. Also, we added new features to the project such that make the user able to recognize the people around him, also recognize currency, and why not to make him recognize color.

Our project is divided into two parts: **Software** and **hardware**

We started with the software which is the coding that programs the raspberry pi to perform these features. We built a class for each feature, then we integrated all classes together.

First: Speech recognition: to execute this feature, we installed

Speech recognition library in python, the function of this class is to take speech and recognize it then convert it to text.

As we support in our module both languages English and Arabic, we built 2 functions "Arabic speech" and "English speech" English speech dealing with all functions in Arabic language and English speech dealing with all functions in English language.

Second: detection of text: we wanted the user to be able hear the text in an image, and to perform this feature, we should first detect the text in the image, and to perform this feature, we used " pytesseract" library in python, because it is the best open-source OCR (Optical Character Recognition) engines which is used to convert typed, printed, or handwritten text into machine-encoded text.

The test input image:

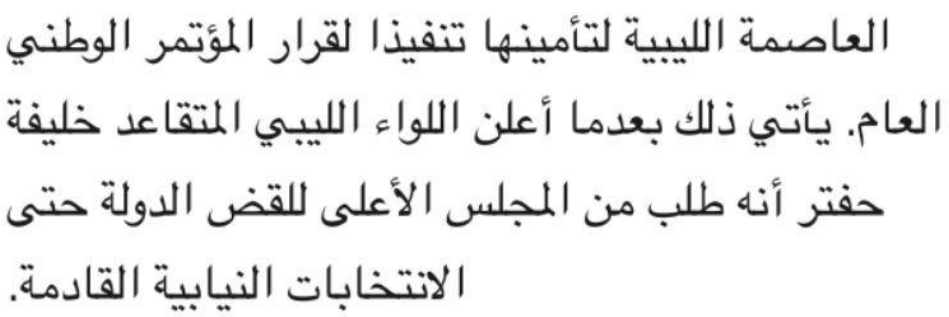

Figure38 input image

And the output on the console that will be converted into audio:

```
العاصمة الليبية لتأمينها تنفيذا لقرار المؤتمر الوطني
العام. يأتي ذلك بعدما أعلن اللواء الليبي المتقاعد خليفة
حفتر أنه طلب من المجلس الأعلى للقض الدولة حتى
الانتخابات النيابية القادمة.
□
[ar:0.999999260351753]
['ar']
ar
```

Figure 39 detected text

Third: as we treat blindness or visually impaired, so we should convert this text into audio message, so they could hear this converted text using headphones, and to perform this feature, we used gtts library in python which stands for google text to speech, to covert the detected text into audio message.

Fourth: it's possible that there are some sentences that user can't understand, and need to translate these sentences, so we used google trans API in python to translate from English into Arabic and vice vera.

for translation from Arabic to English:

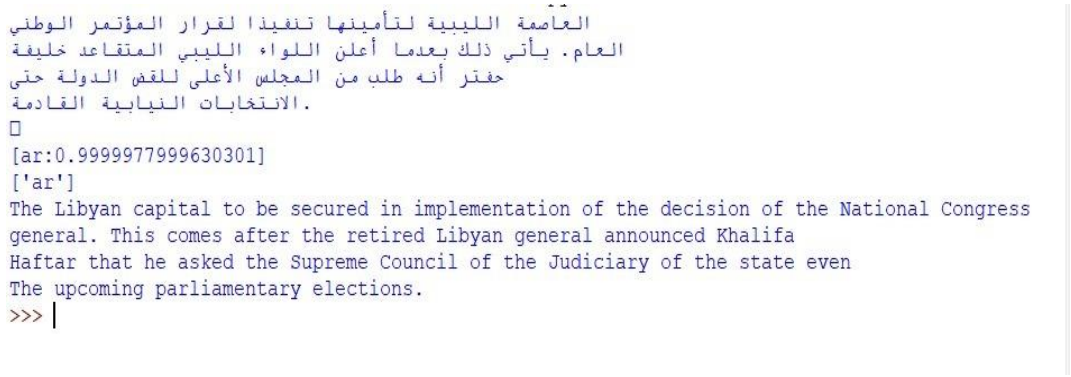

Figure40 translation from Arabic to English

And for translation from English to Arabic:

Figure41 translation from English to Arabic

We notice that there is high accuracy of translation.

Fifth: we wanted the user to be able to recognize the people around him, so we performed face recognition feature using face recognition library in python to allow him to recognize the people around him as his family, friends, or any person he wants, where their images will be stored in the storage of the device as the training images.

It could recognize the person in the input image and said her name “Amal Ashraf Zayed”

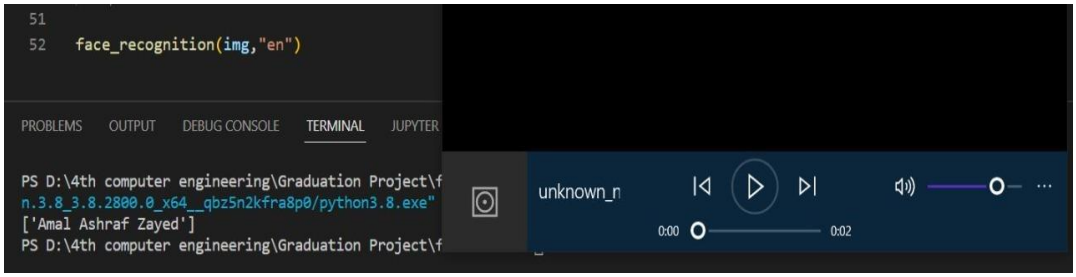

Figure 42

Sixth: we wanted the user to be as independent as possible of himself rather than independent of other people, so we executed the feature that made him recognize currency using ORB (oriented fast and rotated brief) in python, that is used in general for matching between objects and especially we used for matching currency.

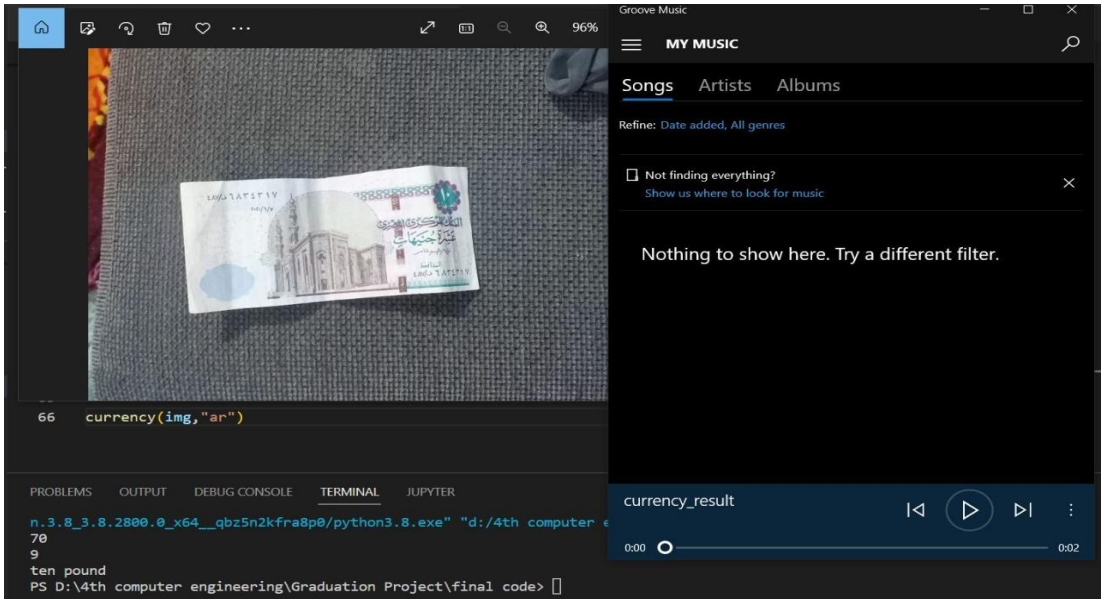

Figure 43

Seventh: we wanted him to be able to recognize the color, so we used OpenCV library in python to execute this feature.

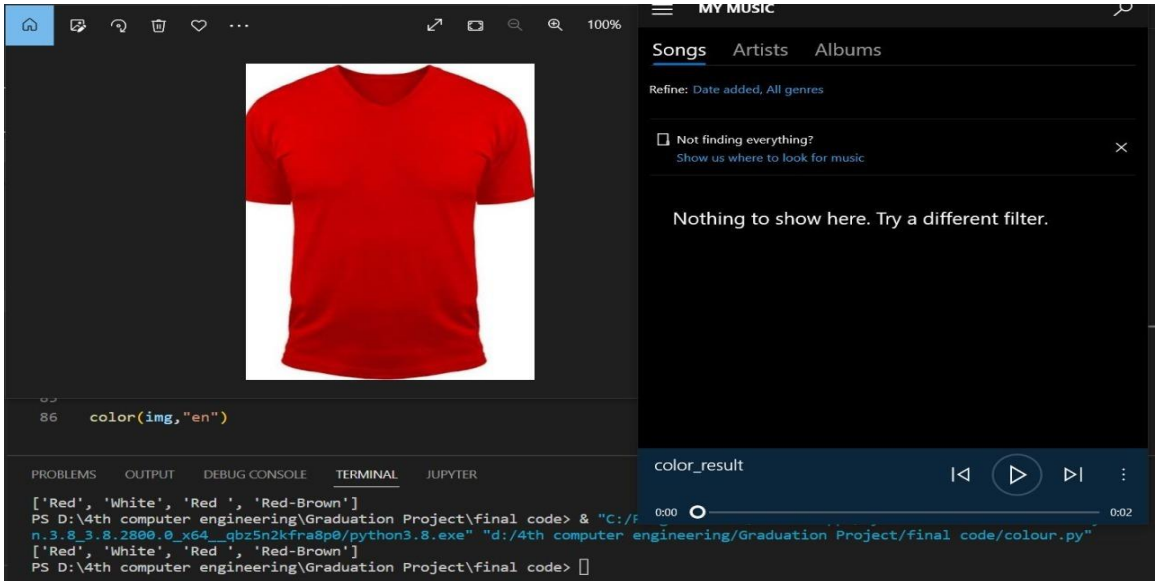

Figure 44

Eighth: we also made it available for user to change the language that treating with the module to perform features from English to Arabic and vice vera, so we built “setting” class to allow the user to make the language as he wants.

The last stage in the software, we integrated all classes together to create a whole module can treat with the user.

After this, we started in the hardware phase, first, we installed the OS on the SD card using: Raspberry pi imager

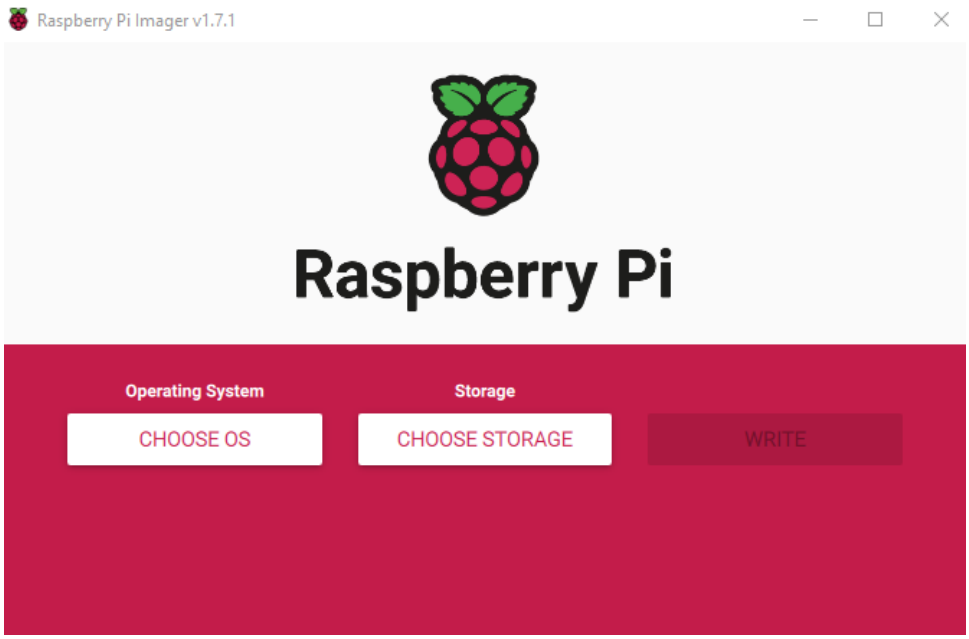

Figure 45 Raspberry pi imager

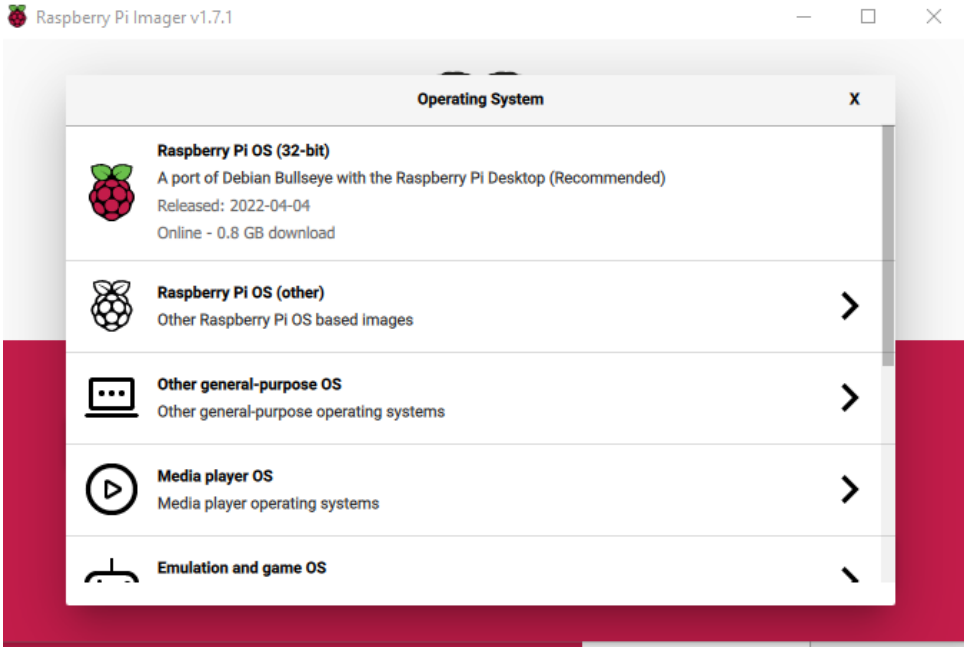

Figure46 selecting OS

We selected the recommended OS to install on the SD card.

Then, we connected the SD card with the raspberry pi, and created username and password, and we burned our python code on the SD card, and successfully executed the functions as expected.

The phase after that was connecting the camera with the raspberry pi and testing its quality and resolution:

As we mentioned earlier, we used webcam with raspberry because of its higher quality.

There are some specifications that are the distance between the image and the webcam should be between 70-90 cm, so that the image is pure, and the module could detect the content of the image whereas is person to recognize or text to detect, currency or color.

The stage after this was to connect the GPS module with the raspberry pi to send location:

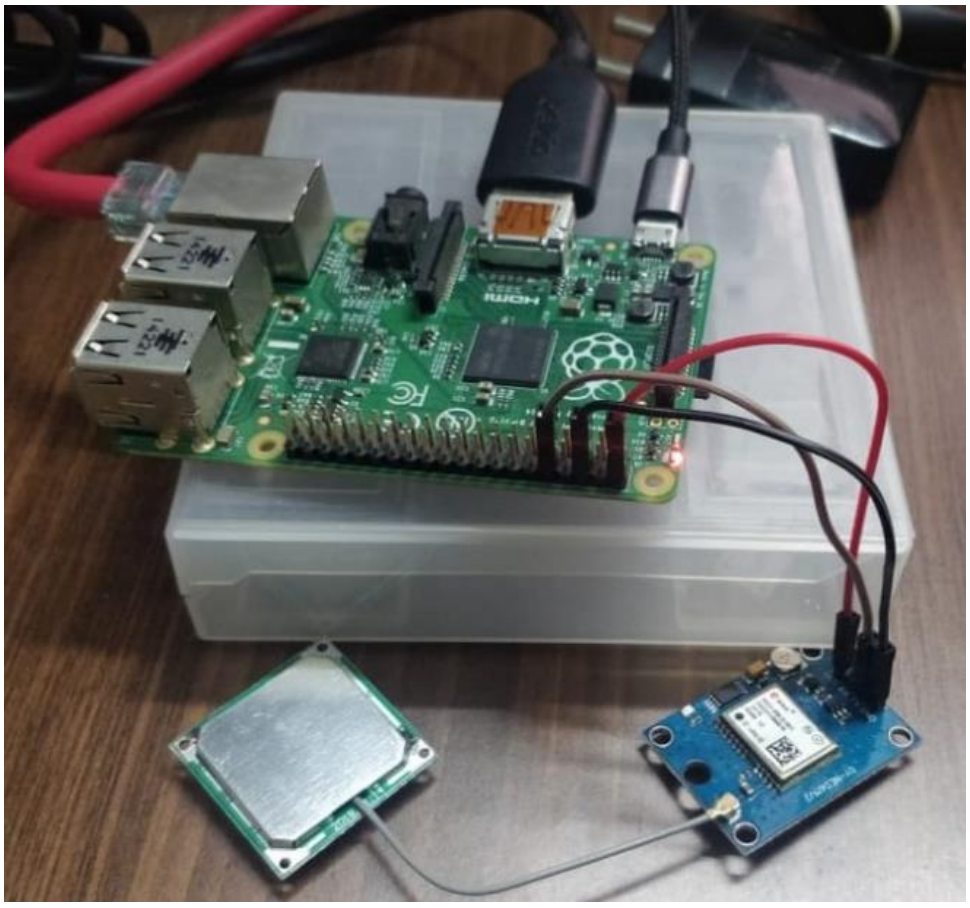

Figure 47

And testing the GPs module (as TX), and it successfully executed as expected, and sent the lat and long to the raspberry pi (as RX), and the raspberry sent SMS message to the phone of the person that the user chose.

This is part of our code to make the GPS module execute its function

```
content ="Look, I'm lost, take me from here ... " +"http://maps.google.com/maps?q=" + str(lat) + "," +
message = client.messages \
        .create(
            body=content,
            from_='[redacted]',
            to='[redacted]'
        )
```

Figure 48

This is the SMS message that is sent to the person the user chose:

And to connect the GPS module with the raspberry pi, we used the pinout configuration of the raspberry Pi

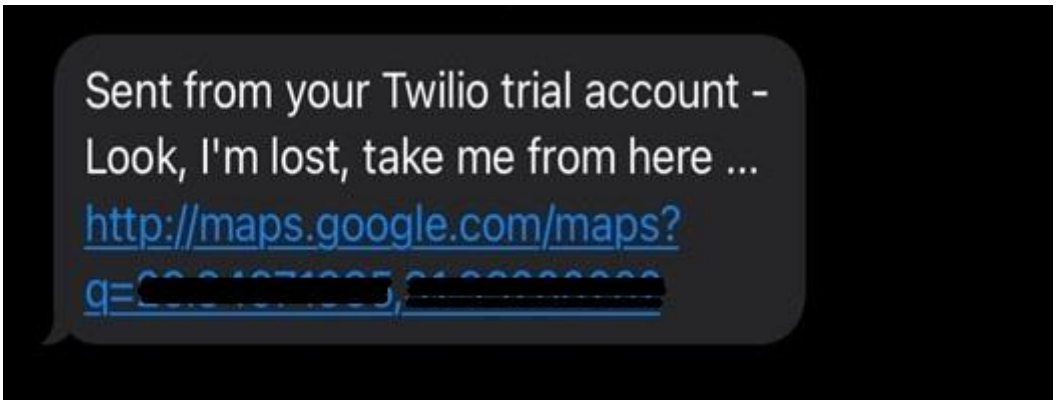

Figure 50 SMS message

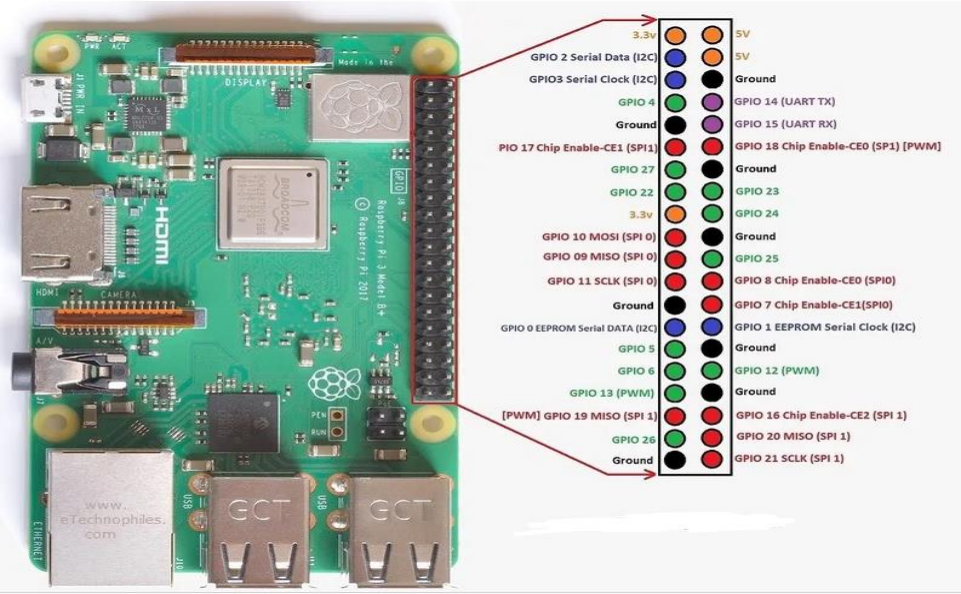

Figure 51 Raspberry pin configuration

The last stage in the hardware was to connect the button with the raspberry pi. we also made it as soon as raspberry connects to power, it will autorun, and when the user presses the button, the camera starts to capture an image.

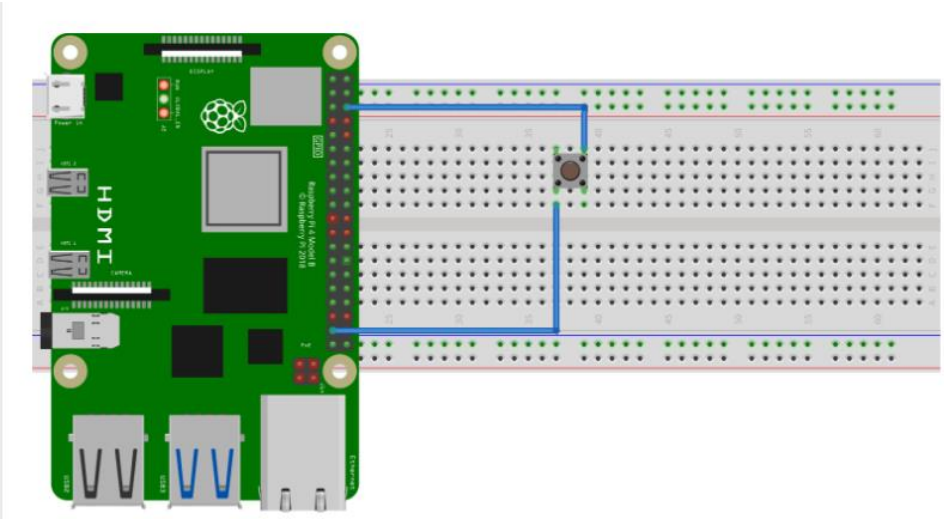

Figure 52 connecting button with raspberry pi

There are some specifications to be performed by the module as:

1- Active Wi fi as all features depends on the input voice of the and the module will respond also with audio messages(gtts) that its function depends on the existence of Wi fi.
2- Also, especially for GPS module, to be able to transmit the location, it should be in an open area (outdoors) to detect the lang and lat of the location and transmit it for the raspberry pi.
3- The distance between the image and the webcam the user wants to detect its contents, so that module could perform the task:
   - between 70-80 cm for face recognition and currency recognition, so that the contents of the image are clear, and the module recognizes

     Correctly.

   - about 40 cm for text detection or text translation so the text is clear and visible for the module so he could perform the task and for color recognition.

Physical:

The team faced some physical constraint during project development. For example, the image that captured by webcam should be clear in order to enable the OCR with OpenCV to detect the text inside it and read it correctly, and also to enable the translation with OpenCV to translate correctly, where the translation of the text depends on the correct detection , to enable face recognition with OpenCV to recognize the people correctly with high accuracy, also The size of the raspberry that is difficult to wear on the head, so the team decides to separate it from the face shield, and we decided that it will be on the arm of the user.

Technical:

To meet the project goal and developed the “Smart virtual assistant” as it should be read and translate the text properly, recognize the people, currency, and color. The team members faced many new technical aspects and learn a new concept, The team learned a new programming language which is python in order to code each component to do its functionality correctly. Also, the team worked on new components such as web cam, raspberry pi module, GPS module, integration of these modules to work together.

Manufacturability:

“Smart virtual assistant” is an extremely helpful device for people with vision difficulties. They can use this product everywhere and improve their lives. “Smart virtual assistant” can be available in different fields like education, entertainment, medical and personal use. There are many people could not be able to continue their studies because they have vision difficulties. So, one of the most important fields is education, by “Smart virtual assistant”, the percentage of educated people will increase. While with “Smart virtual assistant” they could study as a normal person in any school and university.

6. Results and Performance

6.1. Results

we became own a hardware module that helps blind people or people with vision difficulties to read the English or Arabic text in an image , book ,..and also he has the ability to translate any sentence he couldn't understand from English into Arabic or vice versa, to recognize the people around him as his family, friends, ..and also recognize the currency to be able to recognize the amount of money he owns, also recognize the color of t-shirt or any clothes he wants to recognize its color, also the user has the ability to choose the language he wants to treat with the module wheatear it is English or Arabic language. Our module is done to make the blind as possible as independent of the people around them and not all time ,they will need an assistant to accompany them, also these features are implemented so the user could hear the output of any task he selects through the headphones that are connected to the module, also the webcam that is connected to the raspberry module whose function is to capture the image that contains the contents that the user wants to detect wherever is text, face, currency or color, all these tasks are performed under permission of the user when he pressed the button.

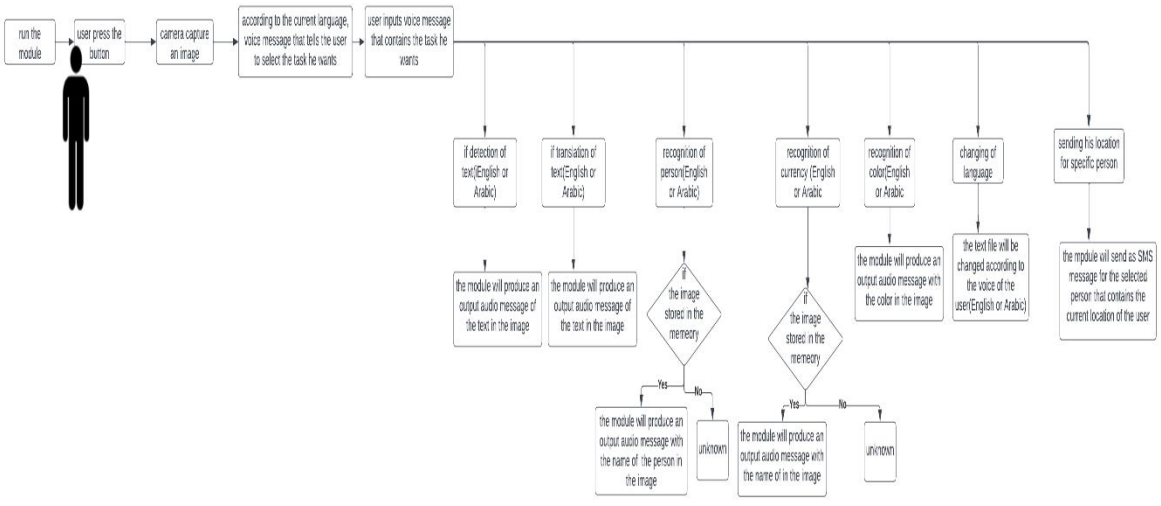

Figure 53 Flowchart of the demonstrated control code

6.2. performance

Feature	Input	Output result	Performance
Text	When the camera is away 40 cm from the text	- The module detects the text correctly - the user listens to the text as a clear audio	**90%**
	When the camera is away 100cm from the text	- The module detects the text less correctly - the user listens to the text as audio	**60%**
translate	When the camera is away 40cm from the text	- The module detects the text correctly and translates it - the user listens to the text after translating as a clear audio	**90%**
	When the camera is away 100cm from the text	- The module detects the text less correctly and fails to translate some of these words - the user listens to the text after translating as audio	**60%**
Face	When the camera is between 70:90 cm from the face	The module could recognize the person correctly, and the user could hear his name through the headphones	**95 %** (This feature is performed with higher accuracy)
Currency	When the camera is between 60:70 from the currency	The module could recognize the currency correctly and the user could hear its amount through the headphones	**90 %** (This feature is performed with higher accuracy)

color	When the camera is away 40cm from the object (T-shirt, jeans,…..)	The module could recognize all colors in the object correctly and the user could hear the colors through the headphones	**85%**
	When the camera is away 80cm from the color (T-shirt , jeans,.)	The module will recognize the colors of the object, but it may mix colors from the background of the image	**75%**
Send location	When the GPS module is indoors	The Gps sends the lat and long=0.0, that means it failed to correctly locate, and the SMS message will be sent to the person with this incorrect lat and long	In general, the performance of the GPS module about 85% (as it needs active Wi-Fi and to be outdoors)
	When the GPS module is outdoors	The GPS module sends the lat and long to the raspberry that will send an SMS message to the person with the current location of the user	

6.3. the final form of module

Figure54 Final module

7. CONCLUSION AND FUTURE SCOPE

7.1. Conclusion

This device aims to be a companion for people with visual disabilities, which allows them to interact naturally with their environment, without the need to acquire or have additional elements

7.2. Future Work

While the team members were working on the implementation, they thought of many ideas and improvements for the “Blind Smart Virtual Assistant”. However, they wished they have more time and knowledge to do them. “Blind Smart Virtual Assistant” can be improved in the future for blind people and people who have vision difficulties by adding new techniques.

For-instance:
-Add commands by which user can control lights, air condition and other home appliances, and alert messages to tell the user about the battery level

- Adding command to adding people in database to recognize them

8. References

1- OTON GLASS | James Dyson Award. (2021). Retrieved 19 December 2021, from https://www.jamesdysonaward.org/en-GB/2016/project/oton-glass/

2- OrCam MyEye 2.0 - For People Who Are Blind or Visually Impaired. (2021). Retrieved 19 December 2021, from https://www.orcam.com/en/myeye2/

3-TechCrunch is part of the Yahoo family of brands. (2021). Retrieved 19 December 2021, from https://techcrunch.com/2018/03/27/airas-new-smart-glasses-give-blind-users-a-guide-throughthe-visual-world/

4-Olguín-Gil L.E., Vázquez-Guzmán F., Vázquez-Zayas E., Mejía J., Blanco-Cruz I. (2022) Virtual Assistant as Support for People Visually Impaired. In: Mejia J., Muñoz M., Rocha Á., Avila-George H., Martínez-Aguilar G.M. (eds) New Perspectives in Software Engineering. CIMPS 2021. Advances in Intelligent Systems and Computing, vol 1416. Springer, Cham. https://doi.org/10.1007/978-3-030-89909-7_14

5- Mori, S., Suen, C. Y., & Yamamoto, K. (1992). Historical review of OCR research and development. Proceedings of the IEEE

6- Thakare, S., Kamble, A., Thengne, V., & Kamble, U. R. (2018). Document Segmentation and Language Translation Using Tesseract-OCR. 2018 IEEE 13th International Conference on Industrial and Information Systems (ICIIS).

7- Bronte, S., Bergasa, L.M., Nuevo, J., Barea, R.: Sistema de Reconocimiento Facial de Conductors (2008). https://tv.uvigo.es/uploads/material/Video/2661/P04.pdf

8- Heo, H. and Lee, K., 2013. FPGA based Implementation of FAST and BRIEF algorithm for Object Recognition. Journal of IKEEE, 17(2), pp.202-207.

9- Minichino, J. and Howse, J., n.d. Learning OpenCV 3 computer vision with Python.

10- (2022). Retrieved 12 July 2022, from https://static.raspberrypi.org/files/product-briefs/Raspberry-Pi-Model-Bplus-Product-Brief.pdf

9. Milestones

<table>
<tr><th>Data</th><th>Milestone</th></tr>
<tr><td>27/12/2021</td><td>Submit Proposal</td></tr>
<tr><td colspan="2"><table>
<tr><td>September/2021</td><td>scan text image and convert it into audio text</td></tr>
<tr><td>September/2021</td><td>translate the text</td></tr>
<tr><td>October/2021</td><td>face recognition</td></tr>
<tr><td>November/2021</td><td>currency recognition</td></tr>
<tr><td>January/2022</td><td>color recognition</td></tr>
<tr><td>February/2022</td><td>Speech Recognition</td></tr>
<tr><td>March/2022</td><td>Integrate classes together</td></tr>
</table></td></tr>
<tr><td>19/3/2022</td><td>Interim Report</td></tr>
<tr><td colspan="2"><table>
<tr><td>April/2022</td><td>Hardware</td></tr>
<tr><td>May /2022</td><td>Sending Location</td></tr>
<tr><td>June/2022</td><td>Finishing project and correct some problems</td></tr>
</table></td></tr>
<tr><td>July/2022</td><td>Final Presentation</td></tr>
</table>