T.C.
ISTANBUL AYDIN UNIVERSITY
INSTİTUTE OF GRADUATE STUDİES

REAL WORLD MAPPING FOR BLIND OR PARTIALLY SIGHTED PEOPLE

MASTER'S THESIS

Imane ASSKOUR

Department of Software Engineering

Artificial Intelligence and Data Science Program

FEBRUARY, 2022

T.C.
ISTANBUL AYDIN UNIVERSITY
INSTİTUTE OF GRADUATE STUDİES

REAL WORLD MAPPING FOR BLIND OR PARTIALLY SIGHTED PEOPLE

MASTER'S THESIS

Imane ASSKOUR

(Y1913.140011)

Department of Software Engineering

Artificial Intelligence and Data Science Program

Thesis Advisor: Prof. Dr. Naim MAHMOOD MUSLEH AJLOUNI

FEBRUARY, 2022

ONAY FORMU

DECLARATION

I hereby declare with respect that the study “Real world Mapping for blind or Partially Sighted People”, which I submitted as a Master thesis, is written without any assistance in violation of scientific ethics and traditions in all the processes from the project phase to the conclusion of the thesis and that the works I have benefited are from those shown in the Bibliography. (06/02/2022).

Imane ASSKOUR

FOREWORD

My first words go to my supervisor Mr Naim AJLOUNI, who has guided me through this research project and didn't hesitate to provide me with valuable recommendations and relevant advices that helped me enormously during the realization of this project.

I am also so grateful for my parents and family for their constant love and support that kept me motivated and confident even in the most difficult times. Any accomplishments and success realized are because of their belief in me.

Finally, special thanks go to the teaching and administrative staffs of the Istanbul Aydın.

February 2022 Imane ASSKOUR

REAL WORLD MAPPING FOR BLİND OR PARTİALLY SİGHTED

ABSTRACT

People who have visual impairment face many problems doing their simple daily activities without needing the help of other people or specially trained dogs. In this study, it is intended to design and develop an optimal solution. The solution will be based on artificial intelligence technologies such as computer vision, intelligent text to voice conversion, detecting objects and/or obstacles, and conveying the information verbally. This will offer modern assistance techniques such as smart glasses and canes, helping the needy to navigate indoor and outdoor.

The solution will also provide a book reading capability. In this study, a system offering indoor and outdoor object detection, obstacle detection and real-time text recognition and conversion to audio will provide users with information to help them overcome their navigation and detection problems. To investigate the ability of the system to achieve its designated task effectively, it has been tested under different environmental conditions including effective performance under high light, darkness the results indicates that the system performs equally well in both cases. The text to voice ability was tested, in which case the system detected video subtitles and text from images equally well.

Keywords: visual impairment, smart glasses, objects detection, computer vision, artificial intelligence, and OCR.

KISMİ YA DA TAMAMEN GÖRME ENGELLİLER İÇİN HARİTALAMA ARACININ TASARIM VE GELİŞTİRİLMESİ

ÖZET

Görme bozukluğu olan kişiler, diğer insanların veya özel olarak eğitilmiş köpeklerin yardımına ihtiyaç duymadan basit günlük aktivitelerini yaparken birçok sorunla karşı karşıya kalırlar. Bu çalışma, optimal bir çözüm tasarlamayı ve geliştirmeyi amaçlamaktadır. Çözüm, bilgisayarla görme, akıllı metinden sese dönüştürme, nesneleri ve/veya engelleri algılama ve bilgiyi sözlü olarak aktarma gibi yapay zeka teknolojilerine dayanacaktır. Bu, akıllı gözlükler ve bastonlar gibi modern yardım teknikleri sunacak ve muhtaçların iç ve dış mekanlarda gezinmesine yardımcı olacaktır. Çözüm ayrıca bir kitap okuma özelliği de sağlayacaktır.

Bu çalışmada, iç ve dış mekan nesne algılama, engel algılama, gerçek zamanlı metin tanıma ve sese dönüştürme sunan bir sistemdir, kullanıcılara navigasyon ve algılama sorunlarının üstesinden gelmelerine yardımcı olacak bilgiler sağlayacaktır. Sistemin belirlenen görevi etkin bir şekilde yerine getirme yeteneğini, normal gün ışığı, yüksek veya parlak ışık ve toplam karanlık altında alet performansı dahil olmak üzere çevresel test koşullarını kontrol etmek için kapsamlı bir test yapılacaktır.

Metinden sese yeteneği test edildi, bu durumda sistem video altyazılarını ve görüntülerden gelen metni eşit derecede iyi tespit etti. Sonuçlar, sistemin her durumda eşit derecede iyi performans gösterdiğini göstermektedir.

Anahtar Kelimeler: Görme bozukluğu, yapay zeka, akıllı gözlükler, nesne algılama, Bilgisayar görüşü.

TABLE OF CONTENTS

LIST OF ABBREVIATIONS

AI : Artificial Intelligence

BVI : Blind and Visually Impaired

CNN : Convolutional Neural Network

ConV : Convolutional

CPU : Central Processing Unit

CSI : Camera Serial Interface

DSP : Digital Signal Processing

FPS : Frame per Second

GPIO : General Purpose Input/output

GPS : Global Positioning System

HDMI : High-Definition Multimedia Interface

HTML : Hypertext Mark-up Language

IR : Infrared

LPDDR : Low Power Double Data Rate

MISO **:** Master in Slave Out

MOSI : Master Out Slave In

MSCOCO **:** Microsoft Common Objects in Context

NLP : Natural Language Processing

OCR : Optical Character Recognition

OpenCV : Open Source Computer Vision

PoE Hat : Power over Ethernet HAT

RCNN : Region-based Convolutional Neural Network

RPi : Raspberry Pi

SD : Secure Digital

SDRAM : Random Access Memory

SR : Speech Recognition

SSD : Single Shot Multibox Detector

SSML : Speech Synthesis Mark-up Language

TTS : Text To Speech

USB : Universal Serial Bus

WHO : World Health Organization

LIST OF TABLES

LIST OF FIGURES

LIST OF EQUATIONS

I. INTRODUCTION

Eyesight is a valuable sensing instrument for humans; it gives humans the ability to investigate their environment. According to the World Health Organization (URL-1), there are about 285 million visually impaired people worldwide (39 million being blind and 246 million having a low vision). Vision's problems in most cases are due to an injury, a disease, age-related degradation. Persons with this handicap face a lot of difficulties while trying to execute their essential day to day needs (navigation, reading books, etc.). For these reasons many assistive devices are commercially available; the most popular are cane guides and dogs. However these classical tools do not provide all the information required such as speed and distance, which are necessary for the perception and the control of locomotion during navigation, also they can't help with reading books or road signs...etc.

Moreover, there are many methods and systems that have been proposed to assist visually impaired people in navigating on a sidewalk (F. Ahmed et al., 2018) or to recognize and read text from objects (Norharyati et al. 2019), etc, but most of the proposed methods are not efficient.

Another wearable, more efficient and comfortable detection tool was smart glasses'. This tool offers the user a clear vision while maintaining a performance similar to a computer. Several companies have worked to develop their version of the smart glasses' including Google in 2013 with a tool called "Google Glass" followed by Eyesynth (URL-2), eSight (URL-3), NuEyes Pro (URL-4), AngleEye (URL-5), and orCam (URL-6). These innovations have in common an embedded system connected to Wi-Fi or Bluetooth to establish wireless communications. These smart glasses have some limitations; the majority offer text reading but not object detection (eSight, Eyesynth, NuEyes Pro) and also, all of them do not support the low light mode.

The proposed project will offer the possibility of recognizing and identifying objects using the raspberry pi infrared module camera, followed by a classification process in real time, and then using the text to speech learning transition to generate

voice feedback. Object detection is achieved in all lighting conditions (dark, low, and bright light).

The system will also have the possibility of measuring object distance and provide the user with suitable warning; the object can be closer or very close, identifying an obstacle as a person or object. The system will also allow reading books or text in real time and converting it into audio. The system also offers the possibility to convert any generated text into one of the three proposed languages (Turkish, English and French).

A. Previous Studies

Many studies and research were carried out to improve the quality of life for visually impaired persons'. The author in (Barontini F et al., 2020) interviewed several blind people to formulate an idea about the features and functionalities that ideal navigation tool should have; the main concept of this system were based on Haptic information Tactile simulation. In this case, the user wears a hand device that contains a motor to help a blind person while moving around; for example, if he moves right, the device will pinch him on his right hand, etc.

In (Mauro Avila et al., 2015) the author developed a navigation system it provides a visually impaired person with the ability to define addresses or rooms. Also, the author in (Tanveer et al. 2015)aim is to assist them in navigating and tracking their location if they are lost using GPS, the system proposes two languages for voice assistance (Bengali and English).

In (Real et al., 2019) the author presented techniques to help future developers interested in working and helping blind people; the author discussed the advantages and disadvantages of each of the proposed methods. So the author of this scientific paper has chosen: indoor positioning, computation offloading, or distributed sensing, etc. the author has taken different approaches that have been proposed in the past in order to explain the strengths, and the limits of each method presented.

Some important key features for an ideal walking assistant have been extracted from (M. M. Islam et al., 2019) after a detailed review made. It includes a low cost, lightweight and straightforward design with reliable interior and exterior coverage. Another study (D. Dakopoulos et al. 2010) was made based on comments from

several blind user groups to extract 14 operational characteristics to describe a good walking assistant.

For the detection of objects, a lot of researchers tried to help avoid obstacles were made (Ali Ali et al., 2017); including the use of drones in this field was set as an option (C. Yi et al., 2013)

The author in (Rahmat Tullah et al., 2020) proposes a blind walking aid that uses ultrasonic sensors to measure the distances between detected objects and the visually impaired person and produces a sound from it buzzer if the distance between them is less than 100cm as a warning. This aid also offers the possibility of detecting if there is a pool of water in front of the user. The smartphone's sensors were used in (M. Ehatisham-ul-Haq et al., 2017) to detect the user's activities such as walking, standing, sitting, running etc. The real-time identification was based on the user's physical activity recognition.

The work presented in (Shifa Shaikh et al., 2020) is a machine learning solution whose main aim is to extract features using YOLO framework trades with object detection. The system also provides text to speech option to describe to the user's the surrounding environment. Another good example is (Shankar Sivan et al., 2016) an assistive device that includes several modules such as object detection, text recognition, and door detection and also have a security system to help detect intrudes.

Based on the potential of computer vision services, a prototype was made in (Hasventhran Baskaranet et al., 2019) to generate a narrative that describes the real world by formulating complete sentences such as "It's a group person reading", "It's a bird sitting on a tree" etc. then those sentences are converted into real-time speech. The last example uses raspberry pi as a processor control for face detection and tracking (P.Velrajkumar et al., 2020) using image processing.

Several researches have taken place in object detection (Bai, S et al., 2018 – S. Pehlivan et al. 2019), each offering different solutions for obstacle detection (R. Radhika et al., 2016 – Sami UR Rahman et al., 2018).

However, the main challenge for the visually impaired problem of identification of text and converting the identified text to voice, including reading books, is still valid. The research carried out in (Roy Shilkrot et al,. 2015) provides a

portable reading device that provides an efficient system with acceptable accuracy using computer vision technology to scan printed text.

The system presented (P. Sabin Prasanna et al., 2018) allows visually impaired persons to read a book in real-time by converting the detected text into audio.

II. PROBLEM STATEMENT

A. Reasons and Problems

Blind people or those who have serious vision problems are daily experiencing many challenges, including moving within the enclosed environment or navigating their way in the outside world and/or reading.

Nowadays, many tools are available on the market to help these people, but so far available solutions either not suitable or they only solve one real problem meaning the person will have to rely on several tools to solve several problems. Hence, the main objective of this work is to present the visually impaired with one device that provides the solution to a number of the major challenges that they face.

The device is a real-time assistance system with a high accuracy. It offers the option of navigation both within confined space and outside open space, including navigation within crowded areas, reading, and translation of the text into three languages (English, French and Turkish).

B. Proposed System

The proposed system will use two branches of artificial intelligence, namely deep learning and computer vision, to achieve object and text detection and identification, plus other options to be discussed in this section. The project is divided into two main parts:

- **Object detection:** It will provide the user with vital information regarding objects with his environment, including object and obstacle detection, and allow the user to avoid them in real-time by delivering detected object details as speech form.

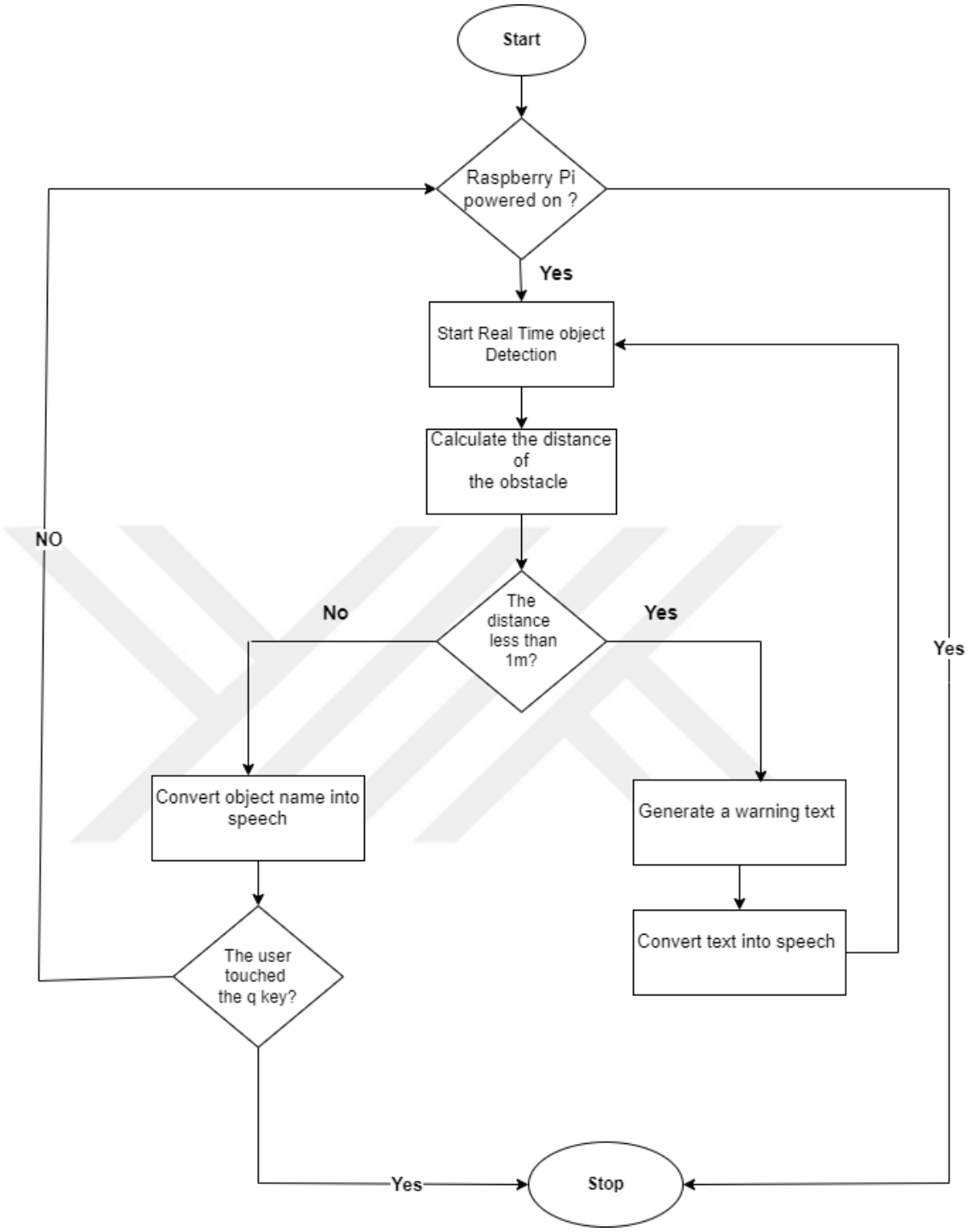

Figure 1 Object Detection flowchart

- **Text reading:** The tool will detect and recognize textual contents within images and or videos; it will convert the text to voice. This means that the tool will allow the user to read and or listen and understand videos containing subtitles. This process will be done by detecting and recognizing the text and then transforming it into speech. The detected text is translated into three languages (Turkish, English and French).

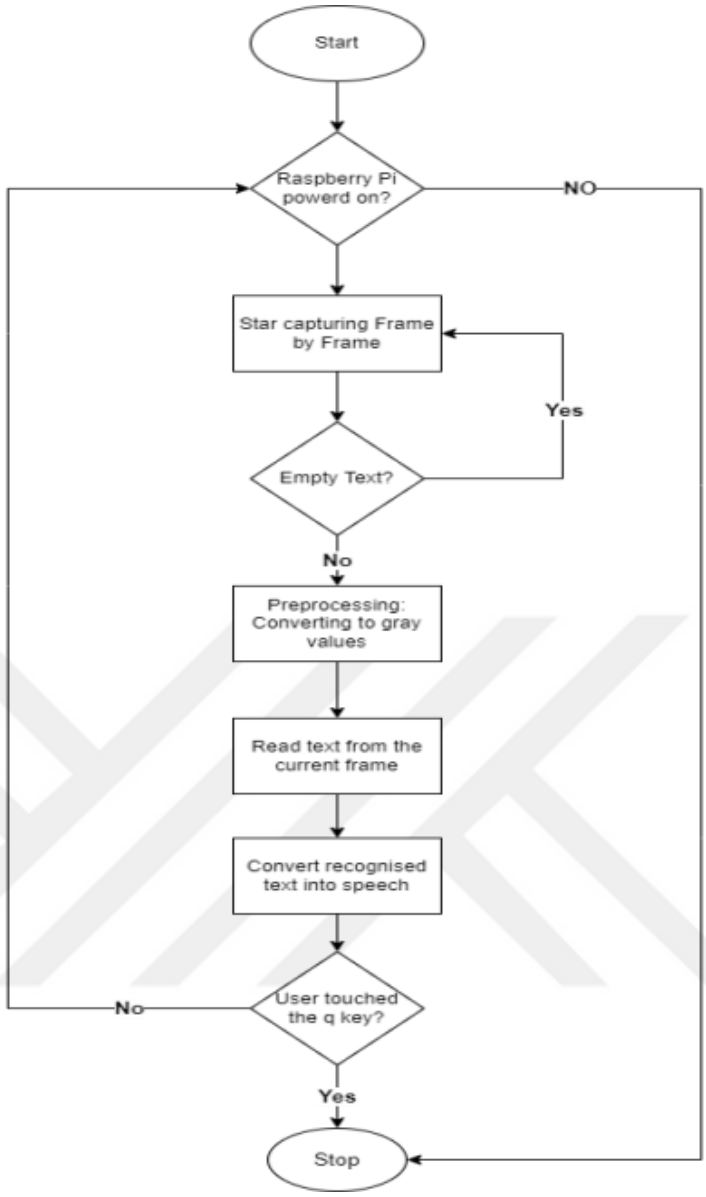

Figure 2 Text Detection Flowchart

Figure 3 illustrates the architecture of the proposed system, which consists of the following modules:

- AI voice menu.
- Object detection.
- Text detection.
- Converting text to speech.
- Distance calculation and obstacle detection.

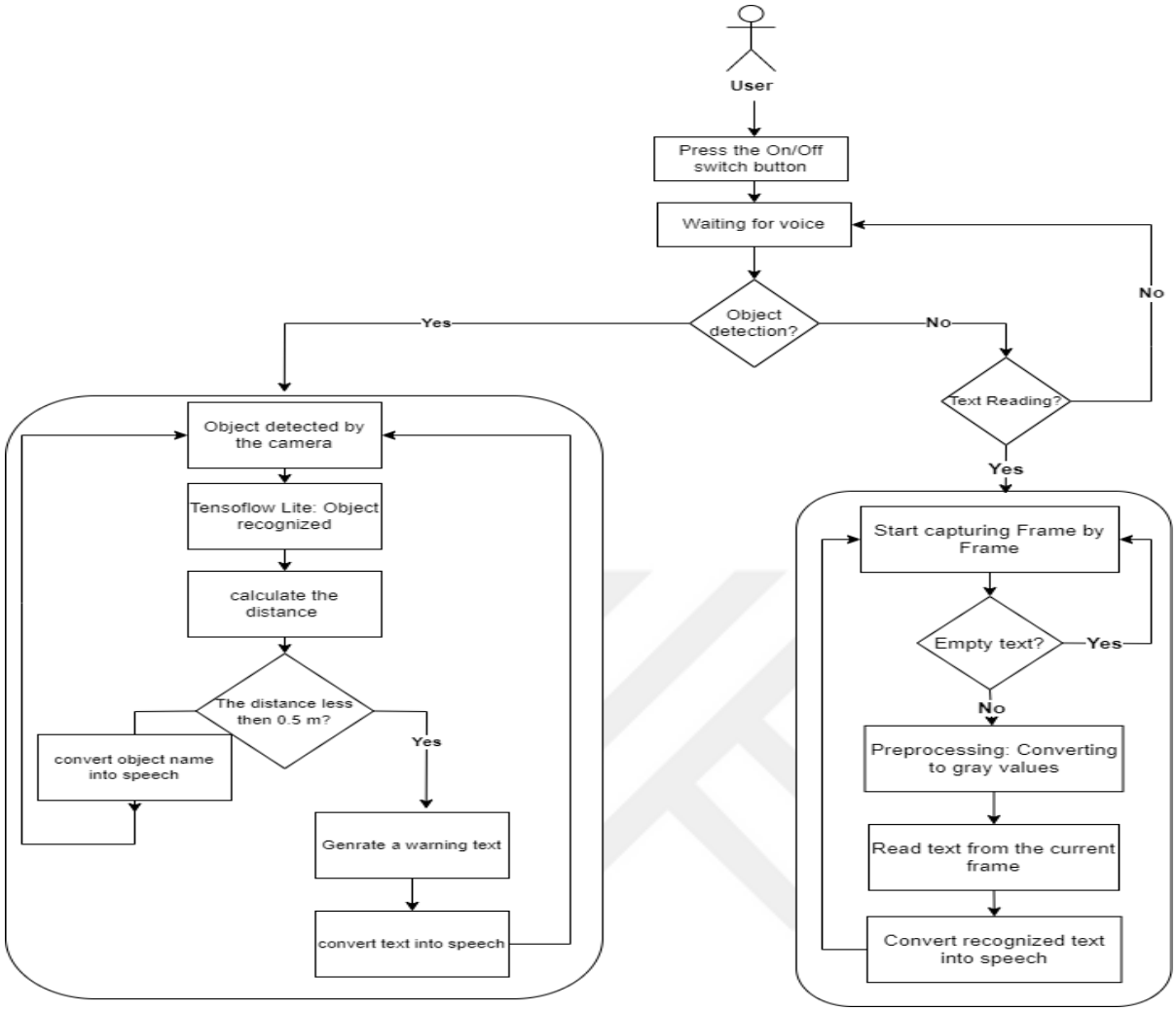

Figure 3 Tool Workflow

III. IMPLEMENTATION

The proposed system consists of hardware and software components; the details of these components are described in this section.

A. Hardware

The hardware components and architecture consists a number of devices and sensors integrated together to achieve the goal of this work. It includes:

1. Raspberry Pi 4

Raspberry Pi is a small computer when connected to a screen and keyboard is used as computer. Its small size and reasonable price make the Raspberry Pi ideal for performing vital tasks, including image and video processing and natural language processing. It also supports several programming languages, including Java, Python, and C ++.

The model used in this project contains 40 pins GPIO; those pins are used to connect the Raspberry pi board to external input/output peripheral devices. Out of these 40 pins, 26 pins are GPIO pins (figure 4).

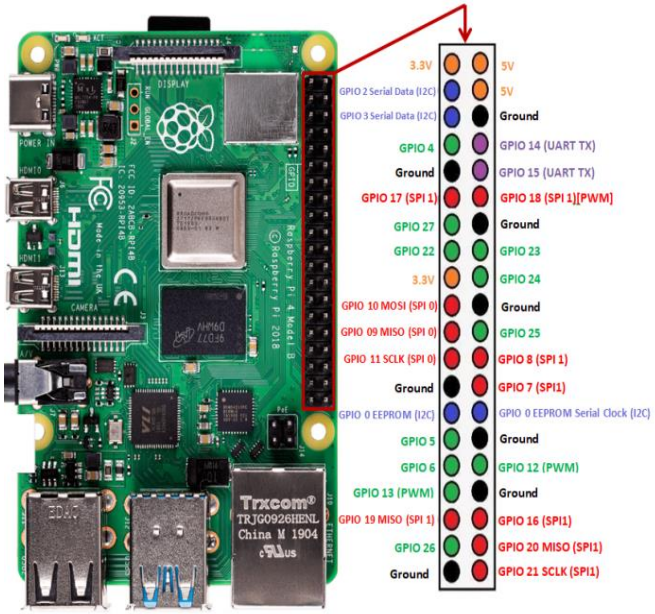

Figure 4 RPi4 GPIO Header

Raspberry pi runs on a Raspbian, which is Debian-based Linux based operating system. There are many models of RPi, but for this project a raspberry pi 4 will act as the central processing unit the system in, this version of RPi offers several major differences compared to the RPi3+, namely more memory, two micro HDMI ports (supporting 4K resolution), etc. (figure4).

2. Ultrasonic Distance Sensor

In this work a number of sensors are used to aid in the achievement of the proposed system task, this includes an ultrasonic sensor. The Ultrasonic sensor is to measure the distance from an object (obstacle) using sound waves. These waves are transmitted at a specific frequency, the sensor in this case waits for the detection of the reflected wave. The distance between the ultrasound sensor and the obstacle is obtained by calculating the time interval between transmission and reception.

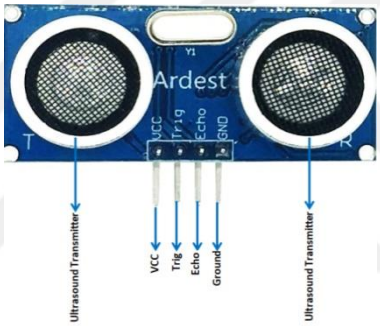

Figure 5 Ultrasonic Distance sensor

A simple ultrasonic sensor consists of one or more ultrasonic emitters, a receiver and a control circuit (figure 5). The output voltage delivered by the Echo bronchus of the HC-SR04 module is 5 volts or the input branch of the RPi is designed for a maximum of 3.3 volts. So in order to avoid damaging the RPi, a voltage divider bridge made up of two resistors is used.

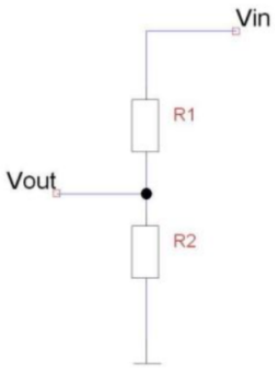

Figure 6 Voltage divider

Using the following formula, we can notice that the adequate resistances that can be used are 1 kΩ and 2 kΩ.

$$\frac{Vout}{Vin} = \frac{R2}{R1+R2} \quad (1)$$

Equation 1 Voltage divider

The ultrasonic sensor may not be able to accurately predict the distance if there are a number of objects at the same time some of the object reflect the wave as good as other objects. However due to coast and financial budget the sensor is used to verify the tools ability to detect and deal with objects and obstacles.

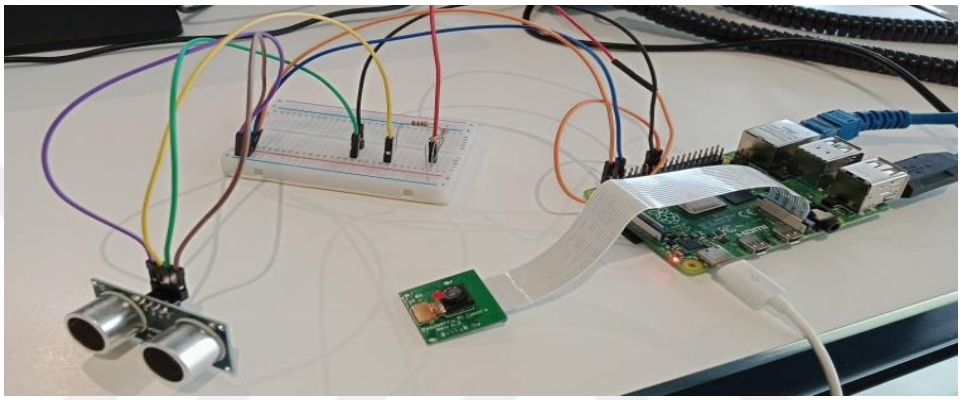

Figure 7 Ultrasonic sensor for obstacle detection

As shown in the figure 7, the connections of the sensor with the RPi are as follow:

– The 5V GPIO (pin 2) is connected to the Vcc pin of the sensor for the 5 V supply;

– The GPIO GND pin is connected to the Gnd pin of the sensor for grounding;

– The GPIO pin 23, output pin of the GPIO is connected to the Trig input of the sensor;

– The GPIO pin 24, GPIO input pin is connected to the Echo output of the sensor via the voltage divider bridge.

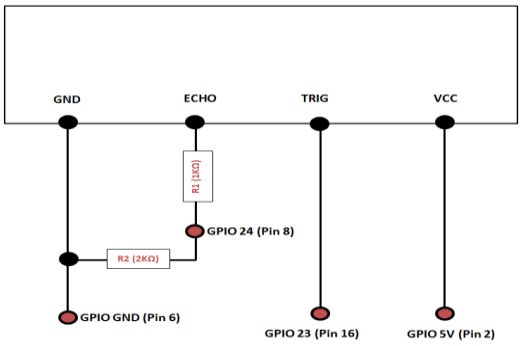

Figure 8 voltage divider bridge

3. Raspberry Pi camera

The following module camera has the possibility to take a high definition video. It also captures images which are passed to the RPi. The image capture frequency is set after testing the tool under varied operation conditions to achieve optimal performance.

Figure 9 Raspberry Pi camera

4. Sound Sensor

Sound sensor is microphone sensor that detects the environmental sound and generates a digital output signal. The sensor is equipped with a potentiometer that allows sense clapping hand including horn sound, sudden and loud voices or environmental noise. The optimal setting is achieved by testing the tools operation under extreme conditions. The reasons for using a sound sensor is to provide the tools the necessary ability of detecting the ambient sound recognition and sound level recognition which will be used to estimate the source of the sounds as well as using the sound waves to identify the identity of the source.

Figure 10 Sound Sensor

So, the sound sensor combines a microphone (50Hz-A0Hz) and a processing circuit, and this to convert the sound waves into electrical signals which are transmitted to the high-precision comparator LM393 in order to digitize them and make them available at the OUT pin.

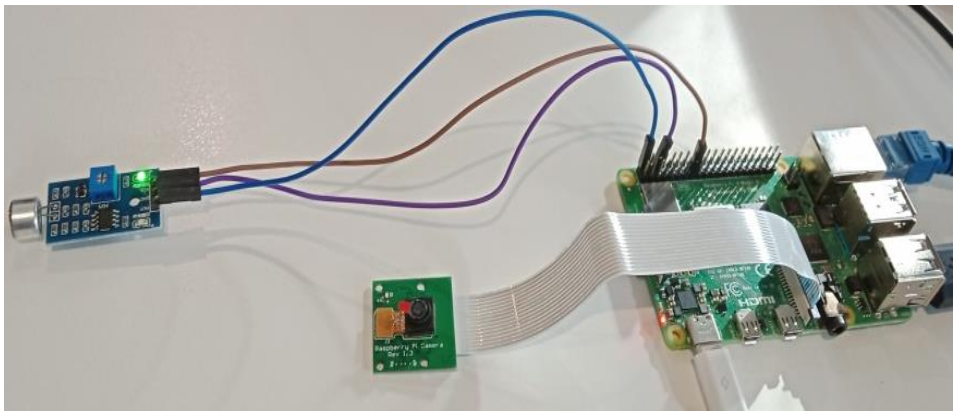

Figure 11 Sound Sensor connection with the RPi

B. Software

A number of software tools are used to develop a working application with the proposed tool. The software tools used include Python is a computer programming language, interactive web tool known as a computational notebook, TensorFlow's lightweight solution for mobile and embedded device, and (Open Source Computer Vision Library).

1. OpenCV

OpenCV (Open Source Computer Vision Library) is a cross-platform library that can help developing real-time computer vision applications. It mainly focuses on image processing; video capture and analysis including features like face detection and object detection.

It was built to provide a common infrastructure for computer vision applications and to accelerate the use of machine perception in the commercial products. Being a BSD-licensed product, OpenCV makes it easy for businesses to utilize and modify the code (URL-7).

In this project the OpenCV is used for the object detection and text detection to grab images and then draw results on them.

2. Speech Recognition

Speech recognition is a ML process in which the sound is passed through some Deep Learning layers and to capture features which can be used to train a model to classification and or prediction. It might be also be used for identifying and or conversion of sound to text. Hence sometimes it can also be called "speech to text".

In this case the machine learning is used to identify words pronounced by a person and convert them into readable text (URL-8). The speech recognition systems process and interpret spoken words via algorithms and then convert them into text. It starts by analysing the audio file; then it is broken into segments and digitized into a computer readable format and finally an appropriate algorithm is used to match it to the most suitable text representation.

3. Optical Character Recognition (OCR)

The technology is used for the live stream text detection and recognition part is the OCR which a technique used to distinguish printed or handwritten text characters inside digital images of physical documents, such as a scanned paper document. The basic process of OCR involves examining the text of a document and translating the characters into digital representation that can be used for data processing. OCR, it is sometimes also referred to as text recognition.

IV. PROPOSED SYSTEM

The system software and related hardware for the proposed tool is discussed in this section. The software components used in this work are designed and developed using the RPi as the processing unit; the related hardware for each of the software components was designed and tuned using the necessary electronic components including potentiometers for achieving optimal hardware signals. Figure 12, presents a block diagram of the proposed system. The components of the proposed system where tested separately and once the final electronic hardware and related software are operating correctly. Once all the components are operational and tested separately they were connected together one component one after another to form the complete system. The integration process included a number of errors which needed in some case modifications to its related hardware and software. Once all the components were connected a complete functionality test was carried out under all possible working conditions to verify the ability of the tool to achieve its goal. From Figure 12, it can been that the system includes a Raspberry pi, Raspberry Pi camera, Ultrasonic sensor, speaker, sound sensor, Three on/off switch unit, and a switch buttons. The system contains five major modules these modules are presented in this section.

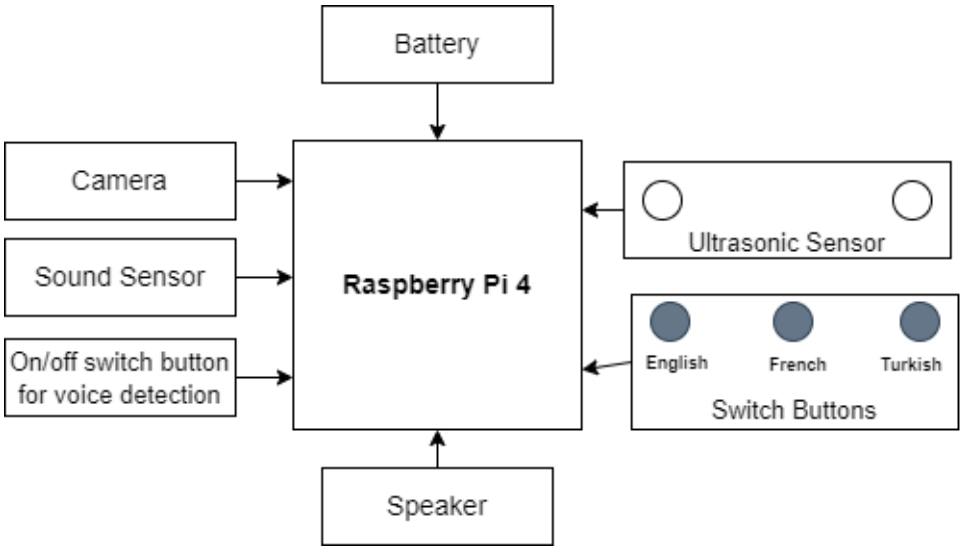

Figure 12 Block diagram

A. Object Detection

The Single Shot Multibox Detector (SDD) is designed for object detection in real time; it takes only one shot to detect multiple objects present in an image using multibox. SSD eliminates bounding box proposals like the ones used in the RCNN's and it includes a progressively decreasing convolutional filter for predicting object categories and offsets in bounding box locations. For those reasons the SSD provides a high accuracy and speed using lower images resolution. SSD architecture (Mark Sandler, et al.)is a CNN framework that consists of two components feature extractor which is also called base network, followed by a set of auxiliary convolutional layers which enable features extraction at multiple scales and decrease the input size of each subsequent layer; and the second component is the bounding box predictor it's a group of small convolutional filters used to predict category scores and box offsets for a fixed set of default bounding boxes.

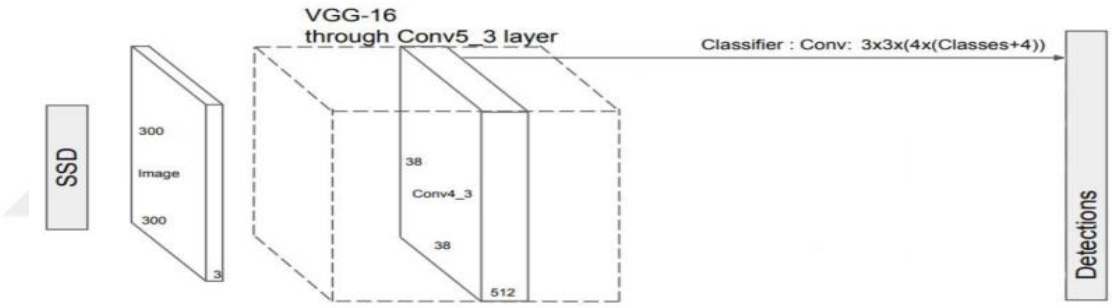

Figure 13 Single Shot multibox Detector

And then the SSD detects objects using ConV layers that evaluate boxes of different aspect ratios at each location in several feature maps with different scales.

1. Convolutional Neural Network

A convolutional neural network (CNN) is a deep artificial neural network that can identify visual patterns from input image with minimal pre-processing compared to other image classification algorithms. The important unit inside a CNN layers is a neuron. They are connected together, in order that the output of neurons at a layer becomes the input of neurons at the next layer.

Figure 14 CNN architecture

It produces highly accurate recognition results and can be retrained for new recognition tasks without the need to build on pre-existing networks.

As shown in figure 13 the CNN model contains 3 types of layers:

- **Convolutional layer:** is the core building block of the CNN. This layer performs a dot product between two matrices, where one is the set of learnable parameters otherwise known as kernel and the other matrix is the restricted portion of the receptive field.
- **Pooling layer:** Replaces the output of the network at certain locations by deriving a summary statistic of the nearby outputs. This helps in reducing the spatial size of the representation, which decreases the required amount of computation and weight. The pooling operation is processed on every slice of the representation individually. There is several pooling function. However the most popular one is the max pooling (figure 14) which reports the maximum output from the neighbourhood.

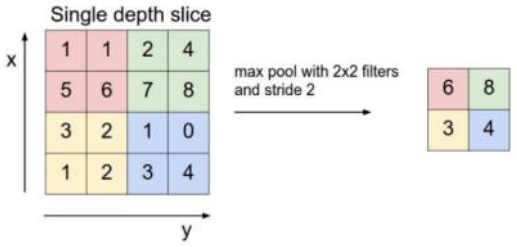

Figure 15 Details on Pooling layer

- **Fully connected layer:** is a traditional Multi-Layer Perceptron that uses an activation function in the output layer to help mapping the representation between the input and the output. The neurons in this layer have full connectivity with all neurons in the preceding and succeeding layer.

There are many types of CNNs. One class of CNNs are depth-wise convolutional neural networks.

2. Depth-Wise

The Depth-Wise is a type of convolution where a single convolutional filter is applied for each input channel. The Depth-Wise convolutions Split the input and filter into channels, convolve each input with the respective filter and finally stack the convolved outputs together (figure 15).

Compared to standard constitutional neural network, the depth wise perform using lesser multiplications; this implies that we can deploy faster convolution neural network models without losing much of the accuracy.

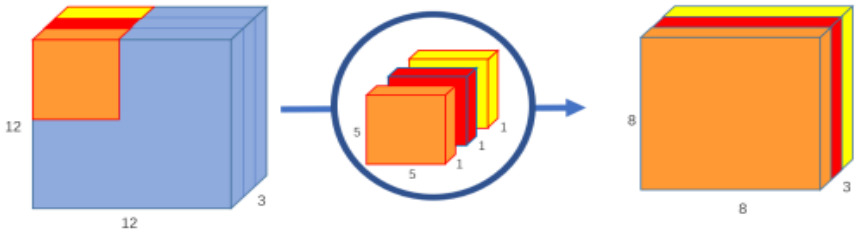

Figure 16 Depthwise convolutions

3. Mobilenetv2-SSDLite

In this work, the SSDLite is used; it is a mobile-friendly variant of SSD, where the regular convolutions in bounding box predictor are replaced by the depth-wise convolutions. Based on the SSDLite framework, the SSDLite-MobileNetV2 (figure 17) utilizes MobileNetV2 as the base network.

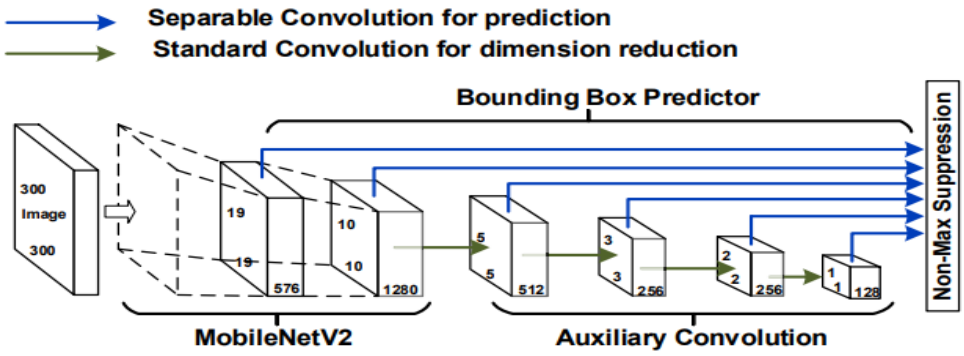

Figure 17 The network architecture of MobileNetV2-SSDLite

So, the model used as an object detector for this system is SSDLite-MobileNet V2 (Mark Sandler et al.) it was trained on the MSCOCO dataset (URL-10) and then converted to run on Tensorflow Lite (URL-11). The MSCOCO dataset (Common Objects in Context) is a large Scale object detection segmentation and captioning dataset published by Microsoft. The format of this dataset is automatically interpreted by advanced neural network libraries.

Some of the COCO datasets features:

- 80 object categories, the “COCO classes” which includes “things” that can be easily labelled (chair, TV, Mobile, cars, cups …etc.)

- 5 captions per image.
- Over 200 000 images of the total 330 000 images are labelled.
- 250 000 people with 17 different key points.

The pseudo code for the object detection algorithm is as follow:

Table 1 Object Detection

Algorithm 1 : Object Detection

```
Initialize video stream
load label map
if label[0] is '???'
    remove label[0]
load the Tensorflow Lite model
while True
    Start timer to calculate frame rate
    Grab frame from video stream
    Acquire frame and resize to expected shape
    Retrieve detection results
    for all detections
        if confidence is above min threshold
            draw detection box
    Draw framerate in corner of frame
    Display the results
    if 'q' button pressed
        break
 end
distroyAllwindows()
stop video stream
```

B. Text Recognition

The Optical Character Recognition (OCR) is the technology used in this system to distinguish printed or handwritten text characters inside digital images of physical documents in real time. The steps involved in the OCR are represented as in figure 18.

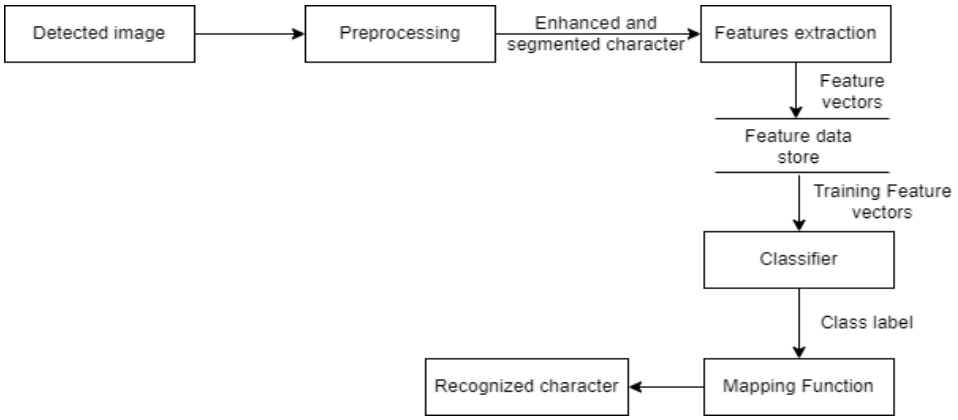

Figure 18 OCR architecture diagram

So, first the raspberry pi camera must be connected and recognized to capture frame by frame, once the current frame is detected, the pre-processing function is executed in order to convert to grey values and this to allow the detection of data even those that are too bright. Then, if the detected frame doesn't show an empty text the data is extracted and a box defining each of it is drawn. The script will continue to be executed while waiting for the user to press the "Q" key and exit the process.

The pseudo code for the real time text recognition is as follow:

Table 2 Real Time Text Detection

Algorithm 2 : Real time text detection
Initialize video stream
while True
Capture frame by frame
Extract any data (text, coordinates, score, etc.)
if box's score is > 60
Extract text
if text **is not** null
start coordinates and dimensions of the box
draw the box
sleep(1)
Convert text to speech
Display the results
if 'q' button pressed
break
end
distroyAllwindows()
stop video stream

C. Text to Speech Conversion

To get audio feedback whether for the object detection part, text recognition or to generate warning messages for obstacle, Text to speech (TTS) synthesizer were used. TTS synthesis means converting written text into understandable natural speech. In this project predefined libraries is used. The structure of TTS is given in Figure 19.

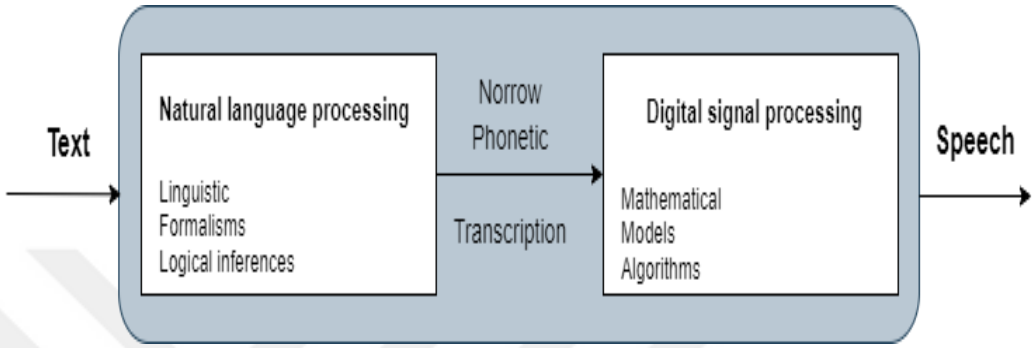

Figure 19 Theory of Text To Speech Synthesis

Natural language processing (Elizabeth D, 2011) involves the reading and understanding of spoken or written language through the medium of a computer, like automatic translation of a language into another, automatic answering of questions..etc.

NLP uses machines to run deep learning algorithms to understand users' text communications and to intelligently respond to them based on the flow configurations.

Digital Signal Processors (DSP) take real-world signals like voice, audio, video, temperature, pressure, or position that has been digitized and then mathematically manipulates them.

A DSP contains these key components:

- **Program Memory**: Stores the programs the DSP will use to process data.
- **Data Memory:** Stores the information to be processed.
- **Compute Engine:** Performs the math processing, accessing the program from the Program Memory and the data from the Data Memory.
- **Input/Output:** Serves a range of functions to connect to the outside world.

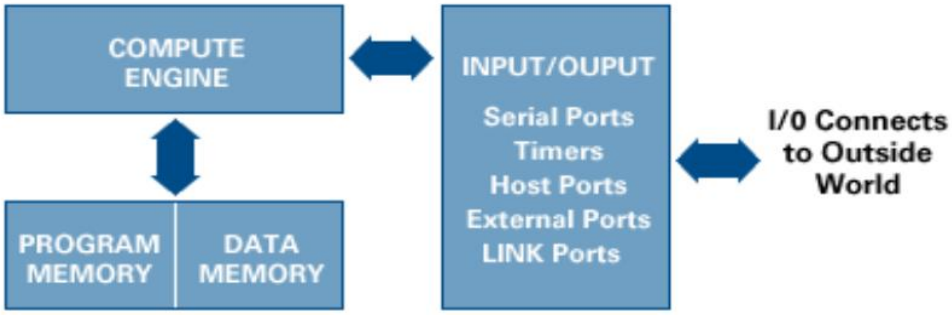

Figure 20 DSP architecture

In this project the Espeak (Ramiz Kastrati et al.) engine is used for conversion of Text to speech. It's a compact open source software text to speech synthesize, it uses a "formant synthesis" method which helped it supports several languages in a small size. The speech proposed by the Espeak engine is clear, and can be used at high speeds.

The Espeak engine features:

– Includes different Voices, whose characteristics can be altered.

– Can produce speech output as a WAV file.

– SSML (Speech Synthesis Mark-up Language) is supported (not complete), and also HTML.

– Compact size. The program and its data, including many languages, totals about few Mbytes.

– Can translate text into phoneme codes, so it could be adapted as a front end for another speech synthesis engine.

– Potential for other languages. Several are included in varying stages of progress. Help from native speakers for these or other languages are welcome.

– Written in C.

D. Obstacle and Distance Measurement

To avoid any obstacle in the forward direction of the device, an Ultrasonic sensor is used. The distance between the device and the obstacle is calculated, if the distance is less than 2m the system generates a warning message which will in turn be converted into audio.

The sensor uses ultrasonic sound to measure distance, it sends out an ultrasonic sound that has a frequency of about 40 kHz. The sensor has two main parts: a transducer that creates an ultrasonic sound and another that listens for its echo.

Sound travels at approximately 340 meters per second. To measure the distance so to measure the distance the following formula is used.

$$\text{Distance} = \frac{(\text{Time} * \text{SpeedOfSound})}{2} \quad (2)$$

Equation 2 Distance calculation

It's divided by 2 because sound has to travel back and forth. The pseudo code for the obstacle detection is as follow:

Table 3 Obstacle Detection

Algorithm 3 : Obstacle detection

```
Set GPIO Pins
Set GPIO direction (IN / OUT)
Try:
    while True:
        calculate the distance
        if distance <= 2 meters
            Generate warning text
            Convert text into speech
            sleep(1)
            draw the box
        sleep(1)
    if 'q' button pressed
        break
end
GPIO clean-up
```

E. Text Translation

To assist users to overcome the language barrier problem, a translation module has been developed. The goal is to give the possibility to the user to choose a language among English, French and Turkish to read a text from an image or a video subtitles. The translation module will be developed based on the Google Translate API.

A switch of three buttons will be used each button represents a language (French, English and Turkish). Once the user presses a button and activates it the text will be translated into the chosen language then converted to speech.

For the hardware part, and to attach the switch to the RPi the following electrical wiring diagram below was used:

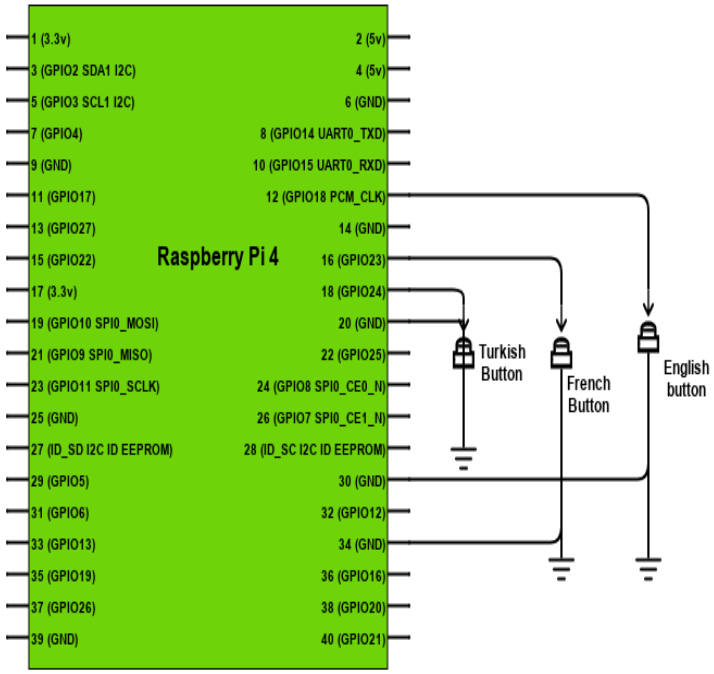

Figure 21 Wiring the switch buttons with Raspberry Pi

When the GPIO pins are set as an input they are “floating” and have no defined voltage level. To be able to reliably detect whether the input is high or low we need to tie it so that it is always connected and either reads high or low. Figure 22 represents the final model after connecting the switch to the RPi.

Figure 22 Switch buttons connection with the RPi

For the text translation module the pseudo code is as follow:

Table 4 Text Translation

Algorithm 4 : Text Translation

```
1: Set GPIO Pins
2: Set GPIO direction (IN / OUT)
3: initialise a previous input variable to 0 (assume the button is not pressed)
4: Try:
6:     while True:
7:         input = GPIO.input(first_button_pin)
8:         if the last reading was low and this one high for the 1st button:
9:             translator = Translator()
10:            Translate text in Turkish
11:            Convert text into speech
12:            Update the previous input button
13:            break
14:        input = GPIO.input(second_button_pin)
15:        if the last reading was low and this one high for the 1st button:
16:            translator = Translator()
17:            Translate text in English
18:            Convert text into speech
19:            Update the previous input button
20:            break
21:        input = GPIO.input(third_button_pin)
22:        if the last reading was low and this one high for the 1st button:
23:            translator = Translator()
24:            Translate text in French
25:            Convert text into speech
26:            Update the previous input button
27:            break
28:    if ‘q’ button pressed
29:        break
30: end
31: GPIO cleanup
```

F. AI Voice Menu

In order to control the device an artificial intelligence voice menu was developed. The objective is to give the possibility to the user to interact verbally with the tool and choose the action he wants to execute.

By default, the object detection module is run, but the user can switch to the real time text reading by giving verbal order via the sound sensor. Once the order is taken by the tool, a warning message asking the user to deactivate the sensor by pressing the switch button will appear.

The work on this module was made based on the Speech Recognition algorithm. Via SR algorithm, any computer software can have the ability to identify words or phrases in spoken language, and then convert them into text.

The block work flow of this algorithm can be summarized as follow:

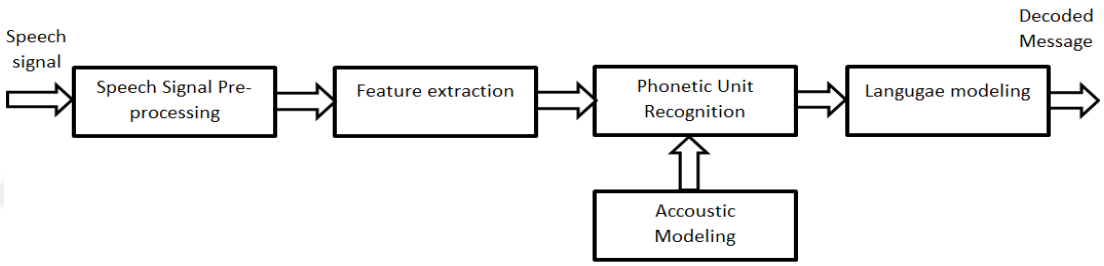

Figure 23 Speech Recognition workflow

Bellow the pseudo code for the artificial intelligence voice menu:

Table 5 AI voice menu

Algorithm 5 : AI voice menu

```
1: Set GPIO Pins
2: Set GPIO direction (IN / OUT)
3: Import Real time object detection code
4: Import Real time text detection code
5: Try:
6:     while True:
7:         input = GPIO.input(button_pin)
8:         if the last reading was low and this one high for the 1st button:
9:                 Create a speech recognition Object
10:                     if GPIO.input(sensor_pin):
11:                      adjust the given input and clear it if there is any noise
12:                      Convert the detected voice input into text
13:                     if detected_text  = “Read”:
14:                         Execute the real time text module
15:                         break
16:                     elif detected_text  = “Navigate”:
17:                         Execute the real time Object detection module
18:                         break
19:                 Update the previous input button
20:                 break
21:     if ‘q’ button pressed
22:         break
23: end
24: GPIO clean-up
```

V. RESULTS & DISCUSSION

To determine the performance of the proposed system, it was tested under various scenarios (Variation in lighting, distance from the camera and number of objects in the frame) and this to identify the object in real-time, recognise text and detect an obstacle, then convert all the obtained results into speech.

A. Normal Light Object Detection

The system was tested in a room with normal lighting; the test results are shown in Figure 24; the system detected the object with high accuracy. The system can also detect more than one object in the same frame and label the objects correctly.

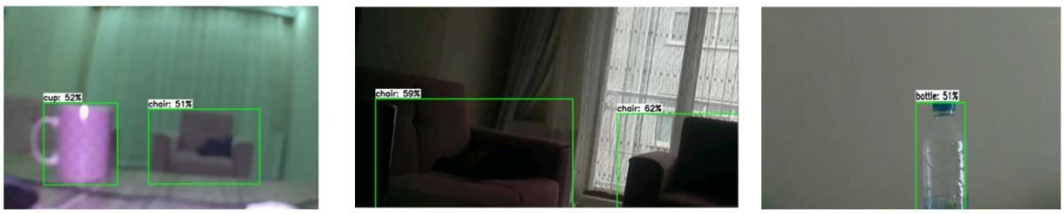

Figure 24 Objects detected in normal light

The present camera had a problem, and it needed to have a magnifying lens attached to it to overcome the problem. Some of the problems and limitations faced during this test include Blair or unrecognised images; this was due to the object's colour and its relation to the surrounding colours, its distance from the camera etc. Figure 25 shows a typical example of an object that the tool failed to recognise. The object is a pink colour armchair and with a white background.

Figure 25 Unrecognized object

C. Text Recognition

The system was tested for reading a scientific printed paper in real-time, as shown in Figure 26, the text was well recognised, and at the same time, it was converted into speech using the Espeak engine.

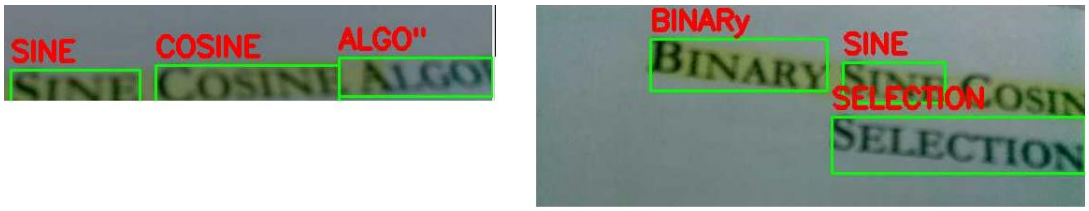

Figure 26 Text recognition

VI. CONCLUSION AND PROPOSALS

An aid tool has been created to help people with visual impairment surround the world and do their daily activities much easier. The system offering indoor and outdoor object detection, obstacle detection and real-time text recognition and conversion to audio and this in order to give the users the information to help them overcome navigation and detection problems. The solution provides also a book reading capability.

The system were tested to check it ability to achieve all the designated tasks, good results were detected after having subjected the hardware to several tests, namely under normal daylight, high or bright light, and total darkness for the object detection part. Also the text to voice ability was tested in the reading of images, video subtitles.

The project can now perform object detection and voice assistance effectively, however, it still need some improvements. So, as a further work a path planning module can be added, to provide a complete and a secure navigation system through an obstacle free optimal path. Connected to Google map, the tool will obtain the most optimal path to follow; and then with the help of verbal assistance, it will give the necessary explanation to the visually impaired person to arrive at his destination and if he needs to take a means of transport, namely the bus or the metro, the module for converting the image into text will be executed to help him read the bus information.

By adding this module, the needy person can be much more independent and won't need any kind of assistance even when his outside and need to go to far destinations.

VII. BIBLIOGRAPHY

ARTICLES

Alberto RODRÍGUEZ, Luis M. BERGASA, "**Obstacle Avoidance System for Assisting Visually Impaired People**", IEEE 2012.

ALI ALI, MOHAMMAD ABOU ALI - **Blind Navigation System for Visually Impaired Using Windowing-Based Mean on Microsoft Kinect Camera** - (ICABME), 2017

BAI, S. LIAN, Z. LIU, K. WANG, and D. LIU, "**Virtual-blind-road following based wearable navigation device for blind people,**" IEEE, Feb. 2018.

BARONTINI F, BETTELANI GC, LEPORINI B, AVERTA G, BIANCHI M. "**A User-Centered Approach to Artificial Sensory Substitution for Blind People Assistance**",2020.

C. YI, R. W. FLORES, R. CHINCHA, and Y. TIAN, "**Finding objects for assisting blind people,**" Netw. Model. Anal. Heal. Informatics Bioinforma, 2013.

D. Dakopoulos and N. G. Bourbakis, "**Wearable obstacle avoidance electronic travel aids for blind: A survey**," IEEE, Jan. 2010

Elizabeth D. LIDDY, "**Natural Language Processing**", Syracuse University, 2011

F. AHMED, S. MAHMUD, R. AL-FAHAD, S. ALAM, and M. YEASIN, "**Image Captioning for Ambient Awareness on a Sidewalk**," 2018

HASVENTHRAN BASKARAN, RACHEL LUM MEI LENG, FIZA ABDUL RAHIM, MOHD EZANEE RUSLI, "**Smart Vision: Assistive Device for the Visually Impaired Community OnlineComputer Vision Service**", IEEE 2019.

HONGXIANG FAN , SHUANGLONG LIU, MARTIN FERIANC, HO-CHEUNG NG, ZHIQIANG QUE, SHEN LIU, XINYU NIU, WAYNE LUK, "**A Real-Time Object Detection Accelerator with Compressed SSDLite on**

FPGA", International Conference on Field-Programmable Technology (FPT) , 2018

I. JOE LOUIS PAUL, S. SASIREKHA, S. MOHANAVALLI, C. JAYASHREE, P. MOOHANAPRIYA and K. MONIKA, "**Smart Eye for Visually Impaired-An aid to help the blind people**,", India, 2019

K. PATIL, Q. JAWADWALA, and F. C. SHU, "**Design and construction of electronic aid for visually impaired people**," IEEE, Apr. 2018.

L. TEPELEA, I. BUCIU, C. GRAVA, I. GAVRILUT and A. GACSÁDI, "**A Vision Module for Visually Impaired People by Using Raspberry PI Platform**,",15th International Conference on Engineering of Modern Electric Systems (EMES), Oradea, Romania, 2019.

M. EHATISHAM-UL-HAQ , MUHAMMAD AWAIS AZAM, USMAN NAEEM , SHAFIQ UR RÈHMAN , ASRA KHALID, **Identifying Smartphone Users based on their Activity Patterns via Mobile Sensing**, The 8th International Conference on Emerging Ubiquitous Systems and Pervasive Networks (EUSPN 2017)

M. M. ISLAM, M. S. SADI, K. Z. ZAMLI, and M. M. AHMED, "**Developing walking assistants for visually impaired people: A review,**" IEEE, Apr. 2019

MARK SANDLER, ANDREW HOWARD, MENGLONG ZHU, ANDREY ZHMOGINOV, LIANG-CHIEH CHEN "**MobileNetV2: Inverted Residuals and Linear Bottlenecks**"

MARK SANDLER, ANDREW HOWARD, MENGLONG ZHU, ANDREY ZHMOGINOV, LIANG-CHIEH CHEN "**MobileNetV2: Inverted Residuals and Linear Bottlenecks**"

MAURO AVILA, MARKUS FUNK, NIELS HENZE – "**DroneNavigator: Using Drones for Navigating Visually Impaired Persons"** - Conference Paper, October 2015

N. RAMA MURTHY, N. SUDHA,"**Smart Navigation System For Visually Challenged People**", International Journal of Industrial Electronics and Electrical Engineering, Sep.-2016.

Norharyati BINTI HARUM, NURUL Azma ZAKARIA, Nurul Akmar EIMRAN, Zakiah AYOP, Syarulnaziah ANAWAR "**Smart Book Reader for Visual Impairment Person using IoT Device**", 2019

P. Sabin PRASANNA, B. Bernadine INFENTA, S. Maria KEERTHANA, Sherril Sophie Maria VINCENT , "**Book Reader using Raspberry Pi**", April 2018

P.VELRAJKUMAR, P.RAMESH, C.SENTHILPARI, T.BHUVANESWARI, V.CHITRA, "**Development of Autonomous Robot for Tunnel Mapping Using Raspberry-Pi Processor**", MARCH 2020

PARTH DHALL, PANKAJ SHARMA, "**A Review Paper on Assistive Shoe & Cane for Visually Impaired People**", International Journal of Scientific Research and Management Studies (IJSRMS), 2015.

R. RADHIKA, P.G. PAI, S. RAKSHITHA and R. SRINATH, "**Implementation of Smart Stick for Obstacle Detection and Navigation**." International Journal of Latest Research in Engineering and Technology, 2016

RAHMAT TULLAH, SYAIPUL RAMDHAN , REZA NABILI AKBAR , FAHMI YUSUF; "**Smart-Cane for The Blind with A Sensor Detection Approach**" Telematika, August (2020)

REAL SANTIAGO, and ALVARO ARAUJO. "**Navigation systems for the blind and visually impaired: Past work, challenges, and open problems**.", (2019)

REBEIRO SHARLENE SARA CARLTON, HUDA NOORDEAN, "**A Review on Object Recognition for Blind People Based on Deep Learning"**, August 2018

ROY SHILKROT, JOCHEN HUBER, WONG MENG EE, PATTIE MAES, and SURANGA CHANDIMA NANAYAKKARA, "**FingerReader: A Wearable Device to Explore Printed Text on the Go**", 2015

S. PEHLIVAN, M. UNAY and A. AKAN, "**Designing Obstacle Detection and Alerting System for Visually Impaired People on Side"**.

SAMI UR RAHMAN, SANA ULLAH, SEHAT ULLAH, "**Obstacle Detection in Indoor Environment for Visually Impaired Using Mobile Camera**", Journal of Physics: Conference Series 2018.

SHANKAR SIVAN, GOPU DARSAN - **Computer Vision based Assistive Technology For Blind and Visually Impaired People**, Conference Paper · July 2016

SHIFA SHAIKH, VRUSHALI KARALE, GAURAV TAWDE, **Assistive Object Recognition System for Visually Impaired**, International Journal of Engineering Research & Technology (IJERT), September-2020.

SREERAJ M, JESTIN JOY, ALPHONSA KURIAKOSE, BHAMEESH M B, ANOOP K BABU, MERIN KUNJUMON, "**VIZIYON: Assistive handled device for visually challenged**", December 2019.

TANVEER, MD SIDDIQUR RAHMAN, M. M. A. HASHEM, AND MD KOWSAR HOSSAIN. "**Android assistant EyeMate for blind and blind tracker**", (ICCIT), IEEE, 2015.

ELECTRONIC SOURCES

URL-1 Blindness and visual impairment, https://www.who.int/news-room/fact-sheets/detail/blindness-and-visual-impairment , (Access Date: 01 February 2022).

URL-2 Inicio - Eyesynth | available on https://eyesynth.com, (Access Date: 01 February 2022).

URL-3 eSight - Electronic eyewear for the visually impaired | available on https://esighteyewear.com , (Access Date: 01 February 2022).

URL-4 NuEyes Pro | available on https://nueyes.com , (Access Date: 01 February 2022).

URL-5 AngelEye Series - AngleEye Smart Reader and AngelEye Smart Glasses | available on https://closingthegap.com (Access Date: 01 February 2022).

URL-6 OrCam MyEye 2.0 - For People Who Are Blind or Visually Impaired | available on https://orcam.com (Access Date: 01 February 2022).

URL-7 OpenCV, https://opencv.org/about/ , (Access Date: 01 February 2022).

URL-8 What is Speech Recognition? | Available on techtarget.com , (Access Date: 01 February 2022).

URL-9 CNN, ujjwalkarn.me/2016/08/11/intuitive-explanation-convnets/ , (Access Date: 01 January 2022)

URL-10 Depthwise Convolution ,https://paperswithcode.com/method/depthwise-convolution/ , (Access Date: 01 January 2022)

URL-11 COCO - Common Objects in Context (cocodataset.org) | https://cocodataset.org , (Access Date: 01 February 2022).

URL-12 TensorFlow Lite | available on https://www.tensorflow.org/lite/guide, (Access Date: 01 February 2022).

APPENDIX

Appendix 1: Raspberry Pi 4

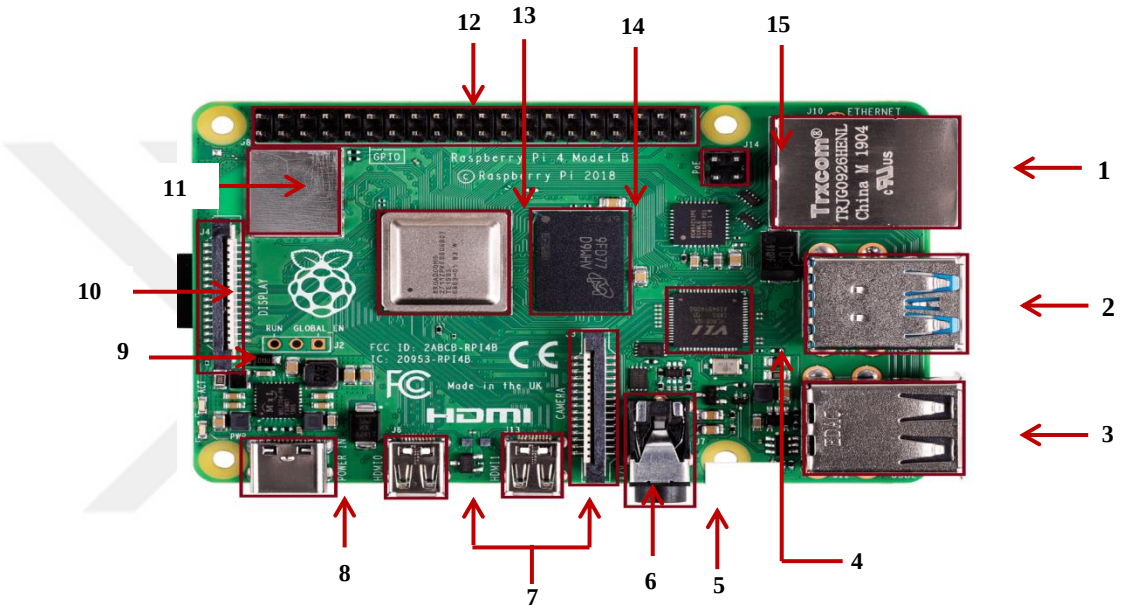

Figure 27 Raspberry Pi 4

1: Ethernet Port it comes with true Gigabit Ethernet capable of sending Ethernet frames at a rate of one gigabit per second (1 billion bits per second).

2: USB port it an USB 3.0 port to connect it to an external keyboard, mouse, or other peripheral devices.

3: USB port it's an USB 2.0 it also helps connecting external peripheral devices.

4: USB Controller

5: Composite Audio-Video Output: Both the audio output socket and the video composite socket reside in a single 4-pole 3.5mm socket.

6: CSI camera port

7: Two micro HDMI ports capable of supporting up to 4k resolution.

8: USB power port, it consists of a 5.1V, 3A USB type-C power port.

9: DSI display port

10: it's a micro-SD card slot (in the back of the raspberry pi) that used for booting up the operating system and storage purposes.

11: Dual band (2.5GHz & 5GHz) Wi-Fi and Bluetooth

12: The model used in this project contains 40 pins GPIO; those pins are used to connect the Raspberry pi board to external input/output peripheral devices. Out of these 40 pins, 26 pins are GPIO pins.

13: CPU it consists of a Broadcom BCM2711 chip which contains a 1.5GHz 64-bit quad-core ARM Cortex-A53 processor (using an ARMv8-architecture core).

14: Low Power Double Data Rate Random Access Memory 4 (LPDDR4 SDRAM) it comes with three different variants 2GB, 4GB and 8GB but the one used in this project is 2GB.

15: PoE Hat Header allows powering the Raspberry Pi using Power over Ethernet–enabled networks.

Appendix 2

Evrak Tarih ve Sayısı: [redacted]

T.C.
İSTANBUL AYDIN ÜNİVERSİTESİ REKTÖRLÜĞÜ
Lisansüstü Eğitim Enstitüsü Müdürlüğü

Sayı : E-[redacted]
Konu : Etik Onayı Hk.

Sayın [redacted]

Tez çalışmanızda kullanmak üzere yapmayı talep ettiğiniz anketiniz İstanbul Aydın Üniversitesi Etik Komisyonu'nun 01.04.2021 tarihli ve 2021/04 sayılı kararıyla uygun bulunmuştur.
Bilgilerinize rica ederim.

Dr.Öğr.Üyesi Alper FİDAN
Müdür Yardımcısı

Dağıtım:
Sayın Dr.Öğr.Üyesi Alper FİDAN
Sayın Ecehan ŞİMŞEK
Sayın Merve AKBULUT
Sayın Gülfidan ERDOĞAN
Sayın Yeşim TİMİSİ
Sayın Buse DURAN
Sayın Feyza ÇAVUŞOĞLU
Sayın Tuğçe DÖNMEZER
Sayın Kadirye Defne DEMİRDÖVEN
Sayın Tuğba SOYSAL
Sayın Özge ALGÜL
Sayın Mayar RANNEH
Sayın Mehmet Ali BALTA
Sayın Ayşe KIZILKUŞ
Sayın Marıam CHAWK
Sayın Danıyal Humaıd Ur REHMAN
Sayın Amır T S AL IDRISI
Sayın Nazir Ahmed NOORI
Sayın Oluwadamilare Dare OSENİ
Sayın Sara Emad Eldin Ahmed Aly
Sayın Awaıs ALSHRAIEDEH
Sayın Hebah Maher Baker HANBALI
Sayın Dmytro BABUKHIN
Sayın RAED KAMHIA
Sayın ZINEB IDDOUTE
Sayın Ayşe Meriç YAZICI
Sayın Sümeyra TİRYAKİ

Bu belge, güvenli elektronik imza ile imzalanmıştır.

Belge Doğrulama Kodu : BSE31F05B3 Pin Kodu : 65032 Belge Takip Adresi : https://www.turkiye.gov.tr/istanbul-aydin-universitesi-ebys?

Adres : Beşyol Mah. İnönü Cad. No:38 Sefaköy , 34295 Küçükçekmece / İSTANBUL
Telefon : 444 1 428
Web : http://www.aydin.edu.tr/
Kep Adresi : iau.yaziisleri@iau.hs03.kep.tr

Bilgi için : Tuğba SÜNNETCİ
Unvanı : Yazı İşleri Uzmanı
Tel No : 31002

CURRICULUM VITAE

Name surname: Imane ASSKOUR

EDUCATION:

- **Bachelor:** 2012, SUP MTI -Rabat-,Information Systems Engineering
- **M.A:** 2014, SUP MTI -Rabat-,, Information Systems Engineering
- **M.A:** 2020, İstanbul Aydin University, Artificial intelligence and data sciences

PROFESSIONAL EXPERIENCES:

- **Computer sciences Teacher:** Elite training and coaching centre, 03/2018 - 01/2020 (Rabat, Morocco)
- **Technical support:** Cegedim Outsourcing Rabat, 04/2016 - 10/2017, (Rabat, Morocco)
- **Intern in software programming:** Poste Maroc, 02/2014 - 06/2014, (Rabat, Morocco)
- **Intern in software programming:** AtoS Rabat, 05/2012 - 06/2012, (Rabat, Morocco)
- **Intern in software programming:** Crédit Agricole Rabat, 09/2011 - 10/2011, (Rabat, Morocco)

PUBLICATION:

ASSKOUR, 2022, Design & Development of Mapping Tool for the Blind or Partially Sighted, Manchester Journal of Artificial Intelligence & Applied Sciences.