

IOT based social distance mask detection and monitoring system

Submitted in partial fulfillment of the requirements for the award of Bachelor of
Engineering degree in Electronics and Communication Engineering

By

38130216 – Surieya M

38130238 – Vignesh T



**DEPARTMENT OF ELECTRONICS AND COMMUNICATION ENGINEERING
SCHOOL OF ELECTRICAL AND ELECTRONICS**

SATHYABAMA INSTITUTE OF SCIENCE AND TECHNOLOGY

(DEEMED TO BE UNIVERSITY)

Accredited with Grade “A” by NAAC

JEPPIAAR NAGAR, RAJIV GANDHI SALAI, CHENNAI - 600119

MAY – 2022

SATHYABAMA

INSTITUTE OF SCIENCE AND TECHNOLOGY

**(DEEMED TO BE UNIVERSITY) Accredited with “A” grade by NAAC I 12B
Status by UGC I Approved by AICTE Jeppiaar Nagar, Rajiv Gandhi Salai ,
Chennai – 600 119**



DEPARTMENT OF ELECTRONICS AND COMMUNICATION ENGINEERING

BONAFIDE CERTIFICATE

This is to certify that this Project Report is the bonafide work of **SURIEYA M** (38130216) and **VIGNESH T** (38130238) who carried out the project entitled **IOT based social distance mask detection and monitoring system** under my supervision from November 2021 to May 2022.

Internal Guide

DR. M. SUMATHI

Head of the Department

DR. RAVI, M.E, Ph.D.,

Submitted for Viva voce Examination held on 05/05/2022

Internal Examiner

External Examiner

DECLARATION

We, **SURIEYA M** (Reg. No.: 38130216) and **VIGNESH T** (Reg. No.: 38130238) hereby declare that the Project Report entitled **IOT based social distance mask detection and monitoring system** by us under the guidance of DR. M SUMATHI is submitted in partial fulfillment of the requirements for the award of Bachelor of Engineering degree in ELECTRONICS AND COMMUNICATION ENGINEERING.

SURIEYA M – 38130216

VIGNESH T –38130238

DATE:

PLACE: CHENNAI

SIGNATURE OF CANDIDATES



Vignesh

ACKNOWLEDGEMENT

I am very pleased to acknowledge my sincere thanks to the **Board of Management of SATHYABAMA** for their kind encouragement in doing this project and for completing it successfully.

We are grateful to them. I convey our thanks to **Dr. N. M. NANDHITHA, M.E., Ph.D.**, Dean, School of Electrical and Electronics and **Dr. RAVI, M.E., Ph.D.**, Head of the Department, Department of Electronics and Communication Engineering for providing me necessary support and details at the right time during the progressive reviews.

I would like to express my sincere and deep sense of gratitude to my Project Guide **DR. M. SUMATHI** for her valuable guidance, suggestions, and constant encouragement paved way for the successful completion of my project work.

I wish to express our thanks to all Teaching and Non-teaching staff members of the Department of Electronics and communication Engineering who were helpful in many ways for the completion of the project.

ABSTRACT:

In this paper, we introduce an affordable IoT-based solution aiming to increase COVID-19 indoor and outdoor safety, covering several relevant aspects mask detection and social distancing check. While mask detection and social distancing check are performed by leveraging computer vision techniques on camera-equipped Raspberry Pi. It helps organizations respect the COVID-19 safety rules and guidelines in order to reduce the disease spread is presented. We focus on most common indoor measures - people with high body temperature should stay at home, wearing mask is obligatory and distance between persons should be at least 1.5-2 meters. For the first scenario, while we rely on Raspberry Pi2 single-board computer equipped with camera making use of computer vision techniques. We decided to use these devices due to their small size and affordability.

LIST OF CONTENTS

CHAPTE R NO	CONTENTS	PAGE NO.
1	Introduction	8-14
2	Literature Review 1. Introduction 2. Navigation Assistance for the Blind. 3. Proposed System. 4. Results and Discussions. 5. Conclusion.	15-22
3	Fundamental Concept of Convolutional Neural Network 1. Introduction. 2. Foundation of Convolutional Neural Network. 3. Concept of Convolution Neural Network.	23-27
4	Proposed System 1. Overview. 2. Exploring the Raspberry Pi Board. 3. Raspberry PI 4 Model B. 4. Hard ware required for Raspberry Pi. 5. Power Supply Unit. 6. Webcam. 7. Image Processing on Raspberry Pi. 8. Introduction of Web Cam Processing. 9. Basic Concept of Image Processing 10. System Hardware Design 11. Ultra Sonic sensing 12. Ultra Sonic Sensor 13. Ultra Sonic - Works 14. Using Multiple Sensors Has Avoiding Description. 15. Ultra Sonic Sensor Used.	28-44

5	Software Description 1. Start up with Raspberry Pi 2. Putty 3. Installation Required Application on Raspberry Pi 4. Python 5. Open CV	45-55
6	Result and Conclusion	56-58

LIST OF FIGURES

FIGURE NO	DESCRIPTION	PAGE NO.
3.1	Extraction layer	26
3.2	The basic conceptual model of CNN	27
4.1	Block diagram	29
4.2	Block diagram of power supply	34
4.3	General block diagram of image processing	40
4.4	System block diagram	41
5.1	Putty login window	46
5.2	Putty success window	47

5.3	Raspberry pi desktop	47
------------	----------------------	-----------

CHAPTER 1

INTRODUCTION

Since the end of 2019, infectious coronavirus disease (COVID-19) has been reported for the first time in Wuhan, and it has become a public damage fitness issue in China and even worldwide. Is pandemic has devastating effects on societies and economies around the world causing a global health crisis. It is an emerging respiratory infectious disease caused by Severe Acute Respiratory Syndrome Coronavirus 2 (SARS-CoV-2). All over the world, especially in the third wave, COVID-19 has been a significant healthcare challenge. Many shutdowns in different industries have been caused by this pandemic. In addition, many sectors such as maintenance projects and infrastructure construction have not been suspended owing to their significant effect on people's routine life.

By now, the virus has rapidly spread to the majority of the countries worldwide. Last statistics (04/05/2021) provided by the World Health Organization (WHO) show 152,543452 confirmed cases and 3198528 deaths. According to the centers for Disease Control and Prevention (CDC), coronavirus infection is transmitted predominantly by respiratory droplets produced when people breathe, talk, cough, or sneeze with common droplet size 5–10'm but aerosol emission increases when humans speak and shout loudly. Pandemic. In addition, many sectors such as maintenance projects and infrastructure construction have not been suspended owing to their significant effect on people's routine life.

Tomography (CT) and X-ray) can lead to a more accurate detection of the COVID-19 and can help to control the problem of loss of specialized physicians in isolated villages .In this context, authors in suggested a novel convolutional neural network (CNN)-based method for detecting COVID-19, with analyzing chest X-ray (CXR) images. Is method allows to detect patients with COVID- 19 at an accuracy of 91.34%. In [18], the authors introduced a new automatic COVID-19 detection model using CXR images. Model is called "DarkCovidNet." For binary classes (COVID-19 VS no findings), the classification accuracy produced by this model is 98.08%, but, for multiclass cases (COVID-19 VS pneumonia VS no findings), the accuracy is 87.02%. Main objective is to use such models to diagnose supplementary chest-related diseases such as tuberculosis and pneumonia. For

the COVID-19 case detection from X-ray images, a deep CNN model is proposed in. Its model, denoted COVID-Net, is open source and accessible to the general public. The detection accuracy achieved by this model (93.3%) proves that the model makes good predictions in improved screening. In [20], distinct deep learning techniques are introduced to differentiate CT scan images of both COVID-19 and non-COVID-19. From the different techniques, we list a self-developed model (CTnet-10), VGG-16, ResNet-50, InceptionV3, VGG-19, and DenseNet-169, which have, respectively, accuracy values of about 82.1%, 89%, 60%, 53.4%, 94.52%, and 93.15%. The accuracy of VGG-19 is the highest one as compared to other models. The CTnet-10 method is a well-organized model, which is useful for doctors, especially in mass screening.

In, two deep learning models are suggested: a CNN and a convolutional long short-term memory (ConvLSTM). To simulate them, two datasets are assumed. A dataset includes

CT images, and the other includes X-ray images. The models are tested four times. When they are examined on CT images, the dataset is split into 70% for the training set and 30% testing set. The accuracy value for the CNN model and for the ConvLSTM is the same, equal to 99%. When tested on the augmented dataset A, the testing accuracy of the CNN is 99%, but it is 100% for the ConvLSTM. When tested on the augmented dataset B, the testing accuracy of the CNN is 100%, but it is 99% for the ConvLSTM. When both models are tested on the combined dataset, containing both X-ray and CT images, the testing accuracy is 99% for the CNN and 98% for the ConvLSTM. Finally, when they are tested on the radiography dataset, the testing accuracy is 95% and 88% for the CNN and the ConvLSTM, respectively. We can consider this scenario as a challenging one, because it is called to distinguish between two diseases (COVID-19 and pneumonia) with a high closeness in features. Before coronavirus, some people put masks to protect themselves from air pollution, while other people put face masks to hide their faces and their emotions from others. Protection against coronavirus is a mandatory counter measure, according to theWHO [1]. Indeed, wearing a mask is an effective method of blocking 80 of all respiratory infections. Also, the WHO recommends practicing physical distancing to mitigate the spread of the virus. All over the world, governments are struggling against this type of virus. Many organizations enforce face mask rules for the personal protection.

Checking manually if individuals entering an organisation are wearing masks is cumbersome and possibly conflicting. In this context, authors in proposed a deep learning-based model, named MobileNet Mask, to prevent human-to-human transmissions of the

SARS-CoV-2 and detect faces with or without mask. Two different datasets (IDS1 and IDS2) with over 5200 images are used to train and test the model. All the experimental cases are controlled on Google Colab that runs in the cloud. In IDS1, the proposed model achieved a testing accuracy of 93%. However, in IDS2, the accuracy achieved is almost 100%.

In [22], the authors aim to detect and delimit medical face masks in real images. The proposed model is composed of YOLO-V2 and ResNet-50. In the training phase, the authors used two optimizers which are Adam and SGDM. During this process, SGDM is better than Adam on validation root mean square error (RMSE), time, and validation loss. However, Adam is better than SGDM in loss and mini batch RMSE. The average precision (AP) of the Adam optimizer is 0.81, which is better than the AP of SGDM that is equal to 0.61. Moreover, the log-average miss rates of the Adam optimizer are 0.4, which are better than the log-average miss rates of the SGDM optimizer (0.6), in all the recall levels. From these studies, we found also, in [23], a system which triggers an alarm in the operating room when the healthcare personnel do not wear face masks. The system combines two detectors, for faces and for masks. In the testing phase, by utilizing images from the BAO dataset and from an own image dataset, above 95% of true positive and below 5% of false positive rates are achieved. Authors in proposed a hybrid model utilizing deep learning with classical machine learning to detect masked faces. The proposed model consists of two components: the feature extraction applying ResNet-50 and the classification process. The three classifiers used are the decision trees (DTs), support vector machine (SVM), and ensemble algorithm. The Real-World Masked Face Dataset (RMFD), the Simulated Masked Face Dataset (SMFD), and the Labeled Faces in the Wild (LFW) are the three face masked datasets, selected for examination. The SVM classifier is greater than the other classifiers. It reached 99.64%, 99.46%, and 100% of testing accuracy, respectively, in RMFD, SMFD, and LFW.

In summary and without forgetting the legal side of AI, the deep learning technique inherently touches upon a full spectrum of legal fields, from legal

philosophy, human rights, contract law, tort law, labor law, criminal law, tax law, procedural law, etc. While in practice, AI is just beginning to come into its own in terms of its use by lawyers, and within the legal industry, legal scholars have been occupied with AI for a long time. Furthermore, as collecting and analyzing data is progressively spreading from software companies to manufacturing companies, which have started to exploit the possibilities arising from collection and exploitation of potential data, so that added value can be created, this information deluge unlocks various legal concerns that could stimulate a regulatory backlash. Considering AI legal concerns and benefits in combating COVID-19 pandemic, AI technique-based solutions are still an open window for development and legal interpretation. The reminder of this paper is organized as follows:

Section 2 summarizes the recent related work in the proposed context.

Section 3 presents the proposed framework. After that, a preliminary study is given in Section 4. Section 5 denotes the dataset collection. Hereafter, we introduce the evaluation metrics in Section 6, while the numerical result is discussed in Section 7. Finally, we conclude this paper in Section 8. Visually challenged people face many problems with accessing printed text, which includes problems with alignment, focus, accuracy, mobility, and efficiency. We present a smart device that assists the visually impaired which effectively and efficiently reads the paper-printed text. The proposed project uses the methodology of a camera based assistive device which will be employed by people to read Text document. The main idea of this project is for implementing image capturing technique in an embedded system based on Raspberry Pi board. In this project, we proposed a text read out a system for the visually challenged. The proposed system uses a camera image as input and the image is processed using OCR software module and the output is given in audio format. A methodology is implemented to recognition sequence of characters, and the line of reading. As a part of the software development, the Open CV (Open source Computer Vision) libraries are utilized to do image capture of text, to do the character recognition. Most of the access technology tools built for people with blindness and limited vision is built on the two basic building blocks of OCR software and Text-to-Speech (TTS) engines. optical character recognition (OCR) is the translation of captured images of printed text into binary codes. OCR may be a process which associates a symbolic meaning with objects(letters, symbols a

number) with the image of a character. It is defined because the process of converting scanned images of machine printed into a computer process able format. An optical Character recognition is additionally useful for visually impaired people that cannot read Text document, but got to access the content of the Text documents. An optical Character recognition is used to digitize and reproduce texts that have been produced with the non-computerized system. Digitizing texts also helps reduce storage space. Editing and Reprinting of a Text document that was printed on paper are time-consuming and labour-intensive. It is widely wont to convert books and documents into electronic files to be used in storage and document analysis. OCR makes it is possible to use techniques like MT, text-to-speech and text mining to the capture/ scanned page. The final recognized text document is fed to the output devices, depending on the choice of the user. The output device are often a headset connected to the raspberry pi or a speaker which may spell out the text document loud.

PURPOSE

- We can build our own **face mask detection model using Raspberry Pi**. This project consists of three phases:
 - Data Gathering
 - Training the model
 - Face Mask Detection
- In the first phase, we will collect the images with and without the mask. This dataset consists of 500 images belonging to two classes.
- In the second phase, we will train the Recognizer for detecting people with and without masks, and in the last phase, we will use the trainer data to classify each face as with a mask or without a mask.
- We previously used Raspberry Pi for some complex image processing projects like facial landmark detection and Face recognition application.

EXISTING SYSTEM

- This paper describes mask detection using Matlab when complex images in the dataset. Matlab specified the Faster R-CNN algorithm and Dataset allotment for mask detection.

- This paper manages complex pictures using facial recognition packages. The Faster R-CNN methodology used in the security system and the medical system. The proposed work balanced face restriction, color changes, brightness changes, and contrast changes.
- Segmentation and feature extraction used in face restriction of the person image. We chose RCNN, Fast RCNN, and Faster RCNN algorithm for detecting Mask detection and Social distance.
- Regions with Convolutional neural network Based on Mixing pictures, pixel prediction, and specific enhancements.

BENEFIT OF PROPOSED SYSTEM

- The system consists of web camera and Raspberry Pi 4. The number of cameras depends on the size of the workspace. The algorithm is programmed to show the path of the person and to display is the social distance was safe or not.
- When algorithm detects a social distancing problem, it alerts and reports the incident to identify frequent hotspot in workplace.
- The algorithm below is consists of several main steps: detection, perspective transformation, measurement and sending e-mail.
- It is Effective.
- It is Low cost.
- We can monitor the blind people with location by receiving the messages.
- Automatic character recognition.

RELATED WORKS

Deep learning is an important breakthrough in the AI field. It has recently shown enormous potential for extracting tiny features in image analysis. Due to the COVID-19 epidemic, some deep learning approaches have been proposed to detect patients infected with coronavirus. In this context, and unlike bacterial pneumonia, many other types of lung infections caused by viruses are called viral pneumonia. These viruses, such as the COVID-19, infect the lungs by blocking the oxygen flow, which can be life-threatening. This motivated researchers to develop many frameworks and schemes based on AI tools in the

fight against this dangerous virus. Hence, we divide this section into two sections to provide an in-depth overview of the proposed techniques.

COVID-19 DETECTION METHODS

Deep Learning Tools and CXR Image-Based COVID-19 Detection. Radiography is a technique used to quantify the functional and structural consequences of chest diseases, to provide high-resolution images on disease progression. Several works have been carried out in this context. Ectiouei et al., in [13], proposed a new CNN-based method for COVID-19 recognition, through analyzing radiographic images of a patient's lungs. The aim of this scheme is to provide clinical decision support for healthcare workers and also for researchers. Hence, performance results, as well as the accuracy value of about 91.34%, and the other metrics in terms of recall, precision, and F1-score, prove the efficiency of the method. In the same context, Ozturk et al., in [18], introduced a new automatic COVID-19 detection model using CXR images denoted by the "DarkCovidNet." It is used to provide correct diagnosis for both a binary classification (COVID-19 VS no findings) and a multiclass classification (COVID-19 VS pneumonia VS no findings). For binary classes, the classification accuracy produced by this model is about 98.08%, but, for multiclass cases, the accuracy is 87.02%. To validate their initial screening, radiologists can use the model to assist them. This model can be employed also via cloud to screen patients immediately. As a solution to the shortage of radiologists, this method can be used in remote places especially in countries affected by COVID-19. The most important advantage of this method is that such models can be used to diagnose supplementary chest-related diseases such as tuberculosis and pneumonia.

CHAPTER 2

LITERATURE REVIEW

This paper presents a prototype to help visually impaired persons in reading printed learning materials using Raspberry PI. Tesseract an open source optical character recognition technique is used extract texts in printed images and converted to audio output using text-to-speech conversion software. Prototype is experimented using printed text pages with various font sizes and line spacing as test cases. Results show that the prototype is better in converting printed texts to speech. However quality of image, font size, and line space affects performance of prototype in converting printed texts to speech..

2. Blind Navigation System Using Image Processing and Embedded System Author Name- Sacinah Jamaludin, Zul Azizi Hailani Description- We got idea from this paper for Navigation system has been developed which help to enhance mobility of blindness. This paper gives us idea for capturing live video of that person & grab video feed in front of blind person & this live video can be observe by admin itself.

2.1 INTRODUCTION

Reading is a part of every human life. Every single day, we read books, magazines, newspapers, etc to update ourselves and to know the developments happens around us [6]. Human eye is an important organ that helps us in this process. But unfortunately visually impaired persons cannot read and update their knowledge as a normal person does. They depend on touching and feeling sense and hearing sense for reading and learning. They need special methods like Braille or Audi lectures for this purpose. learning new things from different sources like books, magazines, newspapers, etc. But unfortunately not everything is printed in Braille. Braille method may not be suitable for quick learning because it supports reading by touching and feeling and requires more time and effort [8]. This makes visually impaired peoples as slow learners. To overcome the above limitation, it is necessary to develop a prototype to that support visually impaired persons in reading printed learning materials..

2. To the best of our knowledge, only very few works are carried out in developing prototypes to help visually impaired peoples in reading printed learning materials. Some of them are discussed below. Ashwini V. Mhaske et al., (2016) proposed,

“Portable Camera Based Assistive Text Reading from Hand Held Objects for Blind Person”. The proposed system assists visually impaired peoples in reading printed texts on hand held objects. A text localization algorithm is proposed using Optical character Recognition (OCR) to perform text recognition on the coverage region and then recognized text codes are transformed to speech. Aaron James S et al (2016) proposed “OCR based automatic book reader for the visually impaired using Raspberry Pi”. In this wok Tesseract algorithm is used for OCR. Texts in scanned images are detected and converted to ASCII codes using Tessact library and open source computer vision library. ASCII codes are converted to audio using filter library and pronounced through earphone using Flite library. Authors discussed about two types of OCR: offline OCR and online OCR. Printed text, handwritten text are used in offline OCR and magnetic ink character recognition are used in online OCR. Prototype is experimented using offline OCR and online OCR need to be explored further. Similar work has been reported by velmurugan D et al (2016) and Mallapa D Gurav eta al., (2017). Ajantha Devi V et al., (2014) proposed “Embedded Optical Character Recognition on Tamil Text Image using Raspberry Pi”. In this work open source computer vision libraries is used recognize tamil text characters and was converted into machine encoded text using OCR. Text-to-speech (TTS) software in raspberry pi converts encoded text in to audio. Authors experimented prototype with Thirukkural text with font size of 10. The challenge in this work is tamil letters has 247 characters out of which 18 are consonants, 216 are compound characters, and 12 vowels. Chucai Yi et al., (2011) proposed "Assistive Text Reading from Complex Background for Blind Persons". In the proposed system authors proposed a camera-based assistive system for visually impaired persons to read text from printed learning materials. The system is able to read text from complex backgrounds and then communicate this information aurally. Performance of the proposed system is evaluated using Robust Reading Dataset and experimental results gives promising solution in reading out texts in printed learning materials. Heba Saleous et al., (2016) proposed “Read2Me: A Cloud-based Reading Aid for the Visually Impaired”. Authors proposed two platforms such as raspberry pi platform and android application platform to help visually impaired people in reading printed text. In raspberry Pi platform text image captured by camera is uploaded by a python web application program in to ABBYY cloud. OCR software

in ABBYY cloud extracts texts in image and sends it back to raspberry pi for further processing. TTS software in raspberry pi converts received text in to audio. A web application program in python is used for communication between raspberry pi and ABBYY cloud. In android application platform text image captured by mobile phone camera is uploaded by a mobile phone app to ABBYY cloud. Texts in image are extracted using OCR techniques in ABBY cloud and send back to mobile phone. Texts are converted into audio using TTS engine in android phones. Experiments are conducted using text images with various font sizes and better results are produced for the font size of 16 and above.

4. Electronic Path Guidance for Visually Impaired People Author Name-Iwan Ulrich and Johann Borenstein Description- From this paper we got idea about the how range which is required for detecting obstacle or object from location of smart stick specific range is there we need to set threshold value if obstacle comes into that threshold range then it can detect successfully otherwise cannot. [4]

5. Use of Ultrasonic Sensors in the Development of an Electronic Travel Aid Author Name- Alex Harold and Chris Gearhart Description- From this paper we got idea about for capturing video image processing is required by making use of some algorithm & method we make some processes on image for capturing it and also live monitoring of that person is seen at admin side. All the processing data is stored on server side in serialized format.[5]

6. Automated Mobility & Orientation System for Blind or Partially Sighted people Author Name- Abdel Ilah_Nour Alshbatat Description- From this paper we got idea about GSM, GPS of sensor like Integrated Ultrasonic sensor, Accelerometer.

2.2 NAVIGATION ASSISTANCE FOR THE BLIND

In today's fast moving world taking each steps without obstacles is a real challenge for the visually impaired. A smart assistant like 3S (Sensory Supervision Spectacles) can assist the visually impaired to read images, avoid obstacles and track routes. So, this paper presents a thought of building up a shrewd framework which can help the visually impaired individuals in their day-to-day life. Some of the main challenges include trouble in moving without the help of someone, difficulty in understanding text or pictures, detecting hindrances, etc. The proposed 3S (Sensory Supervision Spectacles) is capable of text recognition using Optical Character Recognition (OCR), speech synthesis using Text-To-

Speech (TTS), and obstacle detection using ultrasonic sensors (HC-SR04) and a GPS tracker. The segments are associated with Raspberry Pi for being processed. The gadget is a voice empowered framework that would coordinate the visually impaired individual in their everyday works. The estimated number of people visually impaired in the world is 285 million. This number is huge to the point that it drastically affects the economy of the nation. If we look in this busy world, average person doesn't have the opportunity to try and spend some time to assist these distinctively capable ones. In this manner, the visually impaired individuals continually need the help of somebody in their daily life, particularly while on streets and this dependency sometimes leave them to think about themselves in a helplessness or loneliness state. Some other challenges include difficulty in reading signs, symbols, texts or images, detecting obstacles, tracking geographical position, etc. This paper proposes a compact model which can be used everywhere without much difficulty. It incorporates an automated virtual text framework. This is completely processed on a Raspberry. The whole framework is integrated with a camera module, OCR (Optical Character Recognition) mechanical conversion, a TTS (Text-ToSpeech) engine, ultrasonic sensors (HC-SR04) and a GPS (Global Positioning System) satellite-based radio navigation system. The camera module snaps a photo when a user presses the button which is linked with the Raspberry Pi, the saved image is then taken care by the OCR for electronic conversion. OCR accomplishes character recognition via performing some picture handling techniques, (for example, eliminating negative and positive spots). The extracted character will then be saved in a document format (e.g. .txt). This document will then be processed by the TTS software which reads out the text/characters to the user via an output speaker. This model also integrates an object detection module which utilizes ultrasonic sensors which is capable of alerting the user with the exact distance information of the object and the speaker is utilized to produce the alarm sound. Totally four ultrasonic sensors are utilized for covering four directions and alarm sound is produced via output speaker giving four different signals (example: 'left', 'right', 'front', 'back') along with the distance of the object (similar to automobile sensors). Furthermore, the device also gives a vibration signal which aids the user in detecting obstacles in a noisy environment. Hence, the proposed framework is designed in cautioning the user through both voice and vibration. This framework utilizes an Arduino microcontroller, which

gathers all the data from sensor unit and appropriately send signals to response unit to alert the user. The model is additionally integrated with a GPS system, which utilizes internet service for navigation assistance and enables live position tracking by the user's family members.

2.3 PROPOSED SYSTEM

Visually impaired individuals experience various challenges in converting into text utilizing existing technologies, which faces issues with arrangement, centering, precision, portability, and productivity. We present a smart gadget that helps the visually challenged to successfully and effectively reads the printed or written text. The proposed project utilizes the technique of a camera based assistive device that can be utilized by individuals to read printed or written text. The structure is for executing image scanning method in an integrated framework on Raspberry Pi board. In this project, we have proposed a text reading framework for the visually impaired. The proposed completely coordinated framework has a camera as an input device to take care of the printed or written text for digitization, then the document is scanned and processed by a software module called the OCR (Optical Character Recognition engine. Majority of the accessibility tools developed for individuals with visual deficiency and impaired vision are programmed using OCR programming and Text-to-Speech (TTS) engine. OCR makes it conceivable to apply methods, for example, machine interpretation, text-to-speech and text mining to the captured image. And finally the recognized text document is fed to the output device as per user's preference, either a headset or a speaker. This output device reads out the text document loud. The below Fig. 1 shows the basic blocks of the image to Speech Processing Technique

First step of the process is to read the image. This is done with the help of inbuilt camera.

- The next step is preprocessing. Here the image is converted into gray scale image. Then it is converted into black and white image with some threshold value.
- Next the image is cropped into its edges by finding its boundary.
- Once it is finished, the image is converted into word matrix. So that we can

match the letters with templates when template is loaded.

- Each time the extracted letter is compared with the templates. The output is written on a text file.
- Next step is to concatenate. Then it is given to espeak. Thus text is converted into speech.

1) Optical Character Recognition: OCR is an electronic conversion tool which recognizes the text characters in the image which can be in printed text as well as in written form. The OCR layouts of each character are used to recognize the character. Next, the character image is converted into ASCII code, to be used in data processing. ASCII code limits with image recognition, where each character from the image is being picked and sent for recognizing. There are some pre-measures required to make the picture noise free which includes process like binarization, normalization and skew correction. Here we can see the image going through some enhancements, for example, contrast correction and filtering out noise. After these pre-processing, the real role of OCR begins with the segmentation process, in this process each and every characters are separated. This level is assisted by some associated segment examination and projection analysis to again assist with the text segmentation process. Further the sequence is followed by feature extraction. All the real time processing in the OCR lies in the feature extraction.

2) Text to Speech: We realize that the artificial creation of human sound is generally 'speech'. A system is utilized to generate speech known as a speech synthesizer. In that, we can give the sound of human and robot. The text-to-speech framework is divided into two: namely a back-end and front-end. Here, the process begins with the conversion of the raw text like symbols, numbers, abbreviations, etc. This process includes text preprocessing & normalization. Next by allotting the phonetic transcription of the front ends of every word and divides the text into clauses into conditions, prosodic units, expressions, phrases and sentences. The phonetic transcription is assigned into the word. The output of the front end is nothing but the prosody information and phonetic transcriptions of the

symbolic semantic portrayal. Though back-end is transformed into symbolic linguistic portrayal into Audio structure, which is also referred as the synthesizer

2.4 RESULTS AND DISCUSSIONS

The proposed system is tested with different input sets. In this work, the OCR system is implemented for the recognition of English capital alphabets (A to Z) and numbers 0 to 9. The first step is to capture the image. This is the input image which needs to undergo further process. Each character is recognized at one time. The recognized character is saved as text with notepad file, as shown in Fig.9. Next, it converts that text into the speech. The proposed system integrates the working of a HD camera, OCR tool, TTS engine, sensors, speakers and a GPS. All these are integrated using Raspberry pi. The device has been designed in such a way that it is portable and easy for a blind person to use. The real time GPS tracker will assist the blind person to navigate his way to the predefined destination [for e.g. destination set from the home] with the help of the voice assistance feature in the GPS the person can walk with turn-by-turn voice guided navigation feature, on the way if there are any obstacles in the way, the sensors will give signal through voice both direction and the distance of obstacle from the person. This will help the blind person to avoid obstacles in his ways. The blind person can easily capture images with a push of a single button and the text in the images will be read out loud using the proposed system in a short time. As a result, the ultimate goal was to build and design a compact, convenient, and low-cost system that will assist visually impaired people in moving around in unfamiliar environments. The proposed system is built with all ages in mind, is user-friendly, and does not require any prior experience or knowledge of advanced technologies. Additionally, in future development, we can train the system to detect the person present in front of the user. Along with that the system's functionality can be upgraded by training the system with more text font. So that the user can easily detect any text as like a normal person.

2.5 CONCLUSION

Machine learning and artificial intelligence are two of the most rapidly developing technologies. These technologies are critical to the growth of the IT industry. We already have Attempts were made to use these devices for visually impaired people so that they could live independent and normal lives. Blind people who have been denied the gift of sight have the right to live normal lives. In our research, we looked at developing a smart device for blind people to help them live more independently by using digital technology and artificial intelligence.

Chapter 3

FUNDAMENTAL CONCEPTS OF CONVOLUTION NEURAL NETWORK

Convolutional neural network (or CNN) is a special type of multilayer neural network or deep learning architecture inspired by the visual system of living beings. The CNN is very much suitable for different fields of computer vision and natural language processing. The main focus of this chapter is an elaborate discussion of all the basic components of CNN. It also gives a general view of foundation of CNN, recent advancements of CNN and some major application areas.

3.1 Introduction

Among different deep learning architecture, a special type of multilayer neural network for spatial data is Convolutional Neural Network (or CNN or ConvNet.). The architecture of CNN is inspired by the visual perception of living beings. Though it is become popular after the record breaking performance of AlexNet[20] in 2012 but it is actually initiated in 1980. After

2012, the CNN got the pace to take over different fields of computer vision, natural language processing and many more.

The foundation of convolutional neural network started from the discovery of Hubel and Wisel in 1959 [16]. According to them, cells of animal visual cortex recognize light in the small receptive field. In 1980, inspired by this work, Kunihiko Fukusima proposed neocognitron[7].

This network is considered as the first theoretical model for CNN. In 1990, LeCun et al. de-

veloped the modern framework of CNN called LeNet-5[21] to recognize handwritten digits.

Training by backpropagation[35] algorithm helped LeNet-5 in recognizing visual patterns from raw images directly without using any separate feature engineering. But in those days despite of several merits, the performance of CNN in complex problems was lacked by the limited training data, lack of innovation in algorithm and insufficient computing power. Recently we have large labeled datasets, innovative algorithms and powerful GPU machines. In 2012, with these up-gradation, a large deep CNN, called AlexNet, designed by Krizhevsky et al[20]

showed excellent performance on the ILSVRC[36]. The success of AlexNet paved the way to invent

different CNN models[41] as well as to apply those models in different fields of computer vision and natural language processing.

A traditional convolutional neural network is made up of single or multiple blocks of convolu-

tion and pooling layers, followed by one or multiple fully connected (FC) layers and an output layer. The convolutional layer is the core building block of a CNN.

This layer aims to learn feature representations of the input. The convolutional layer is composed of several learnable convolution kernels or filters which are used to compute different feature maps. Each unit of feature map is connected to a receptive field in the previous layer. The new feature map is produced by convolving the input with the kernels and applying elementwise non-linear activation function on the convolved result. The parameter sharing property of convolutional layer reduces the model complexity. Pooling or sub-sampling layer takes a small region of the con-

volutional output as input and downsamples it to produce a single output. There are different sub-sampling techniques as example max pooling, min pooling, average pooling, etc. Pooling reduces the number of parameters to be computed as well as it makes the network translation

invariant. Last part of CNN is basically made up of one or more FC layers typically found in

feedforward neural network. The FC layer takes input from the final pooling or convolutional layer and generates final output of CNN. In case of image classification, a CNN can be viewed as a combination of two parts: feature extraction part and classification part. Both convolu-

tion and pooling layers perform feature extraction. As an example of dog's image, different convolution layers from lower level to higher level detect various features such as two eyes, long ears, four legs, etc for further recognition. On top of this features, the FC layers are added as

classifier, and a probability is assigned for the input image being a dog. Besides the layer design, the improvement of CNN depends on several different aspects such as activation function, normalization method, loss function, regularization, optimization and processing speed, etc.

After the success of AlexNet, CNN got huge popularity in three major fields namely image classification[41], object detection[42] and segmentation[45], and many advance models of CNN has been proposed in those areas in successive years. Major application areas that apply CNN to achieve state-of-the-art performance includes image classification, object tracking, object detection, segmentation, human pose estimation, text detection, visual saliency detection, action recognition, scene labelling, visual question answering, speech and natural language processing, etc. Though current CNN models work very well for various applications, it is yet to know why and how it works essentially. So, more efforts on investigating the fundamental principles of

CNNs are required.

3.2 Foundation of Convolutional Neural Network

In 1959, two neurophysiologists David Hubel and Torsten Wiesel experimented and later published their paper, entitled of single neurons in cat's striate cortex" described that the neurons inside the brain of a cat are organized in layered form. These layers learn how to recognize visual patterns by first extracting the local features and then combining the extracted features for higher level representation. Later on, this concept is essentially become one of the core principle of Deep Learning.

Inspired by the work of Hubel and Wiesel, in 1980, Kunihiro Fukushima proposed Neocognitron, which is a self-organizing Neural Network, containing multiple layers, capable of recognizing visual patterns hierarchically through learning and this architecture became the

first theoretical model of CNN as in the figure 1. A major improvement over the architecture

Of Neocognitron was done by LeCun et. In 1989 by developing a modern framework of CNN, called LeNet-5, which successfully recognized the MNIST handwritten digits dataset. LeNet-5 was trained using error back-propagation algorithm and it can be recognize visual patterns directly from raw input images, without using any separated feature engineering mechanism.

After discovering LeNet-5, because of several limitation like lack of large training data, lack of innovation in algorithm and inadequate computing power, CNN did

not performs well in various complex problems. But nowadays, in the era of Big Data we have large labeled datasets, more innovative algorithms and especially powerfull GPU machines. With these type of up-gradation, in 2012, Krizhevsky et al. designed AexNet, which achieved a fantastic accuracy on the ImageNet Large Scale Visual Recognition Challenge (ILSVRC). The victory of AlexNet paved the way to invent several CNN models as well as to apply those models in di_erent _eld of computer vision and natural language processing.

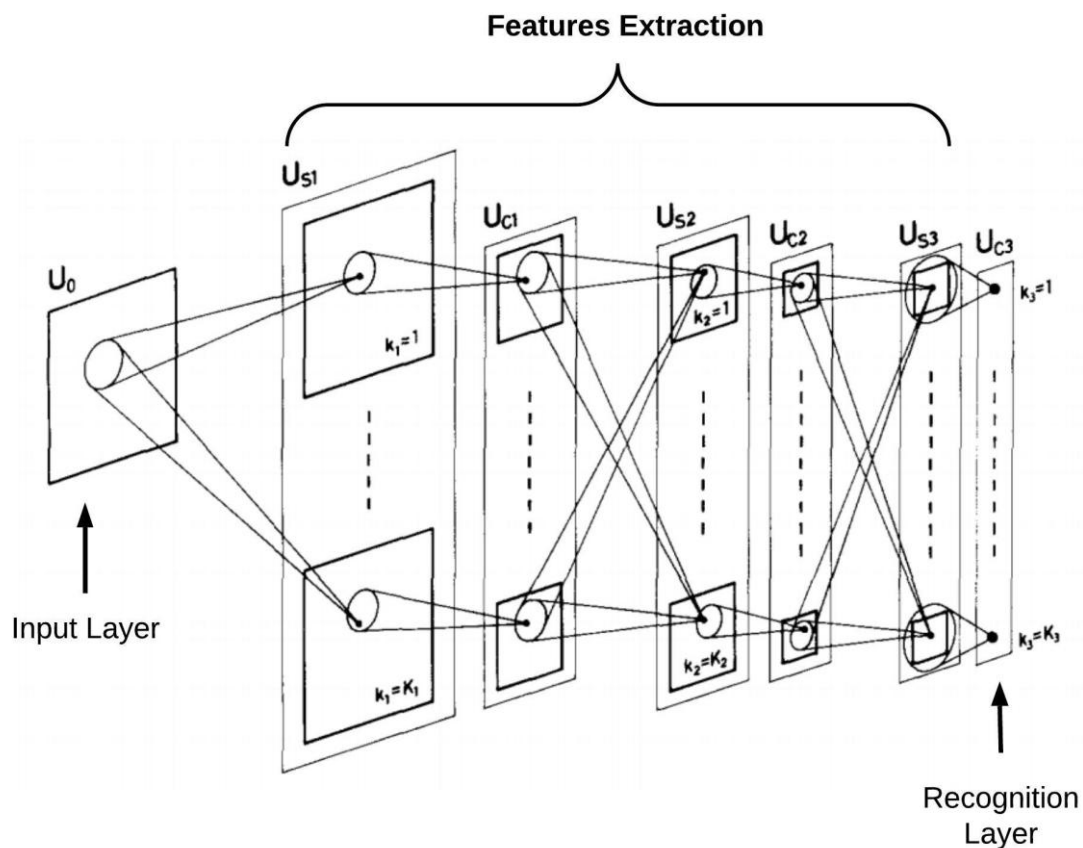


Fig: 3.1

3.3 Concepts of Convolutional Neural Network

Convolutional Neural Network (CNN), also called ConvNet, is a type of Arti_cial Neural Network (ANN), which has deep feed-forward architecture and has amazing generalizing ability as compared to other networks with FC layers, it can learn highly abstracted features of objects especially spatial data and can indentify them more e_ciently.

A deep CNN model consists of a finite set of processing layers that can learn various features of input data (e.g. image) with multiple level of abstraction. The initiatory layers learn and extract the high level features (with lower abstraction), and the deeper layers learns and extracts the low level features (with higher abstraction). The basic conceptual model of CNN was shown in figure 2, different types of layers described in subsequent sections.

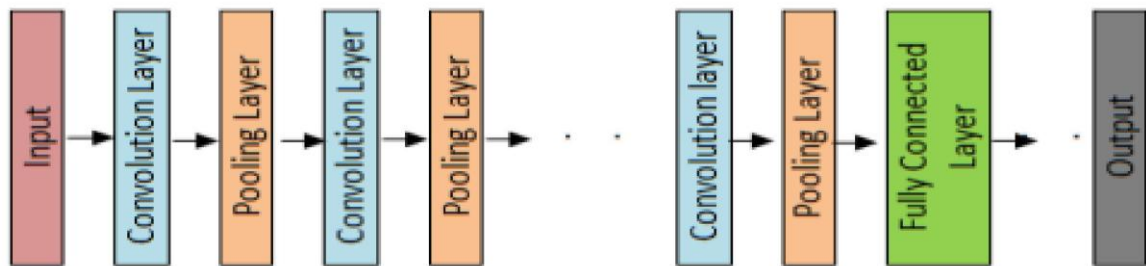


Fig: 3.2

Conceptual model of CNN

- One of the main reason for considering CNN in such case is the weight sharing feature of
- CNN, that reduce the number of trainable parameters in the network, which helped the model to avoid overwriting and as well as to improved generalization.
- In CNN, the classification layer and the feature extraction layers learn together, that makes the output of the model more organized and makes the output more dependent to the extracted features.
- The implementation of a large network is more dicult by using other types of neural networks rather than using Convolutional Neural Networks.

CHAPTER 4

PROPOSED SYSTEM

INTRODUCTION

Therefore, this research paper focuses on implementing a Face Mask and Social Distancing Detection model as an embedded vision system. The pretrained models such as the MobileNet, ResNet Classifier, and VGG are used in our context. People violating social distancing or not wearing masks were detected. After implementing and deploying the models, the selected one achieved a confidence score of 100%. This paper also provides a comparative study of different face detection and face mask classification models. The system performance is evaluated in terms of precision, recall, F1-score, support, sensitivity, specificity, and accuracy that demonstrate the practical applicability. The system performs with F1-score of 99%, sensitivity of 99%, specificity of 99%, and an accuracy of 100%. Hence, this solution tracks the people with or without masks in a real-time scenario and ensures social distancing by generating an alarm if there is a violation in the scene or in public places. This can be used with the existing embedded camera infrastructure to enable these analytics which can be applied to various verticals, as well as in an office building or at airport terminals/gates.

4.1 BLOCK DIAGRAM:

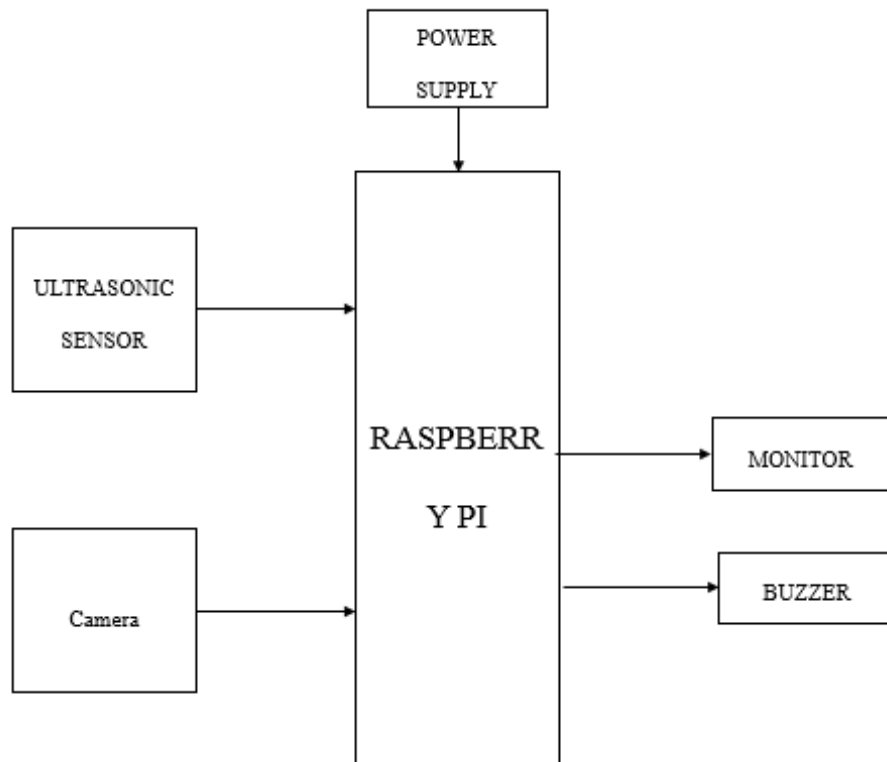


Fig:4.1

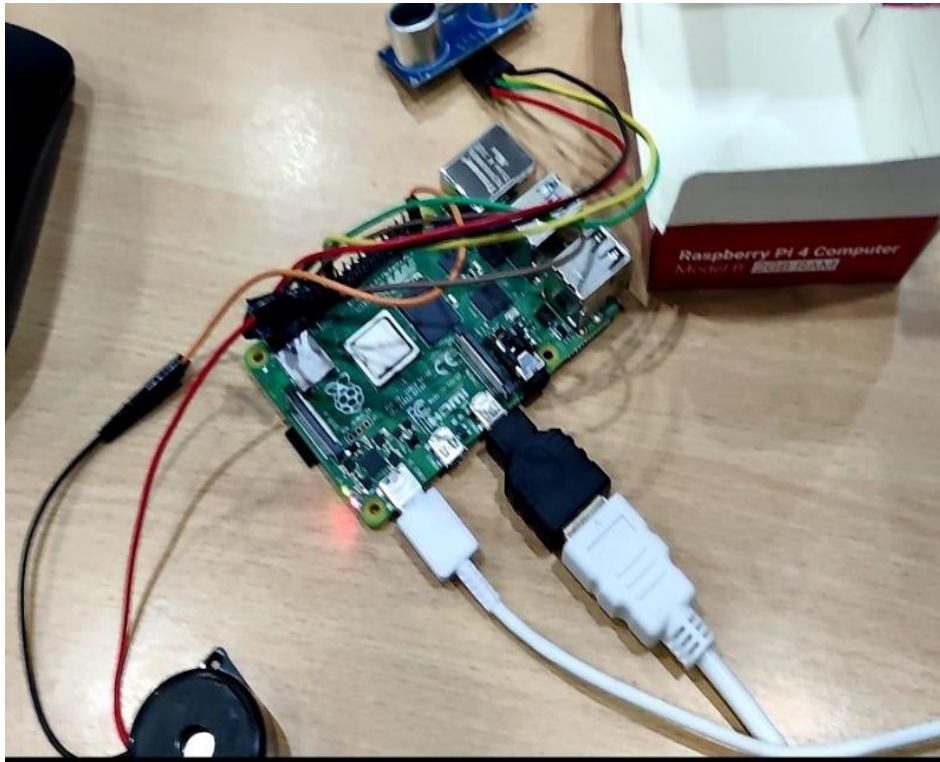
Hardware

- Raspberry Pi
- Web Camera
- Speaker
- Ultrasonic Sensor

Software

- Raspian OS
- Python
- Open CV

Project kit



HARDWARE DESCRIPTION

RASPBERRY PI

4.1 Overview

Raspberry Pi seems to be new in the world and many people really don't know what the Raspberry Pi is. Raspberry Pi can be defined as a low cost, credit-card sized computer that plugs into a computer monitor or TV, and uses a standard keyboard and mouse. It is a capable little device that enables people of all ages to explore computing, and to learn how to program in languages like Scratch and Python. It's capable of doing everything you would expect a desktop computer to do, from browsing the internet and playing high definition video, to making spreadsheets, word-processing, and playing games. It is great bonding with Arduino and can do a lot with Arduino.

4.2 Exploring the Raspberry Pi Board

There are two models of Raspberry Pi, model A and model B. These two are bit similar with few advance features on model B compared to model A. Model B has 512 MB RAM, two USB port where as Model A has 256 MB RAM and just a USB port. Besides, Model B has Ethernet port while Model A does not.

Different components of the Pi are briefly described on each component is given in following sections.

4.2.1 SD Card Slot

Raspberry Pi doesn't have the real hard drive as in laptop and computer, SD card is taken as solid state drive (SSD) which is used to install operating system and all others software and store everything. This card is needed to insert into the slot for using the Raspberry Pi.

SD card may be 2GB, 4GB or 16GB.

4.2.2 Micro USB Power

The power port is a 5V micro-USB input and supply should be exactly 5v as it doesn't have onboard power regulator. So, power supply shouldn't exceed than 5V.

4.2.3 HDMI Out

This output port is used to connect the Raspberry Pi with a monitor via HDMI (High Definition Multimedia Interface). Hence, any screen or TV can be connected to it which consists of HDMI port.

4.2.4 Ethernet and USB port

Both the Ethernet port and USB port on Model B are supplied via the onboard LAN9512 chip. It is a high-speed USB 2.0 hub with a 10/100 Ethernet controller (Donat, 2014). USB ports are used to connect the inputs (keyboard, mouse). Almost everything that can connect to computer via USB also can connect with Raspberry Pi.

4.2.5 RCA Video Out and Audio Out

Audio and RCA video jacks are present on the board for audio and video out.

The Raspberry Pi does support sound over its HDMI output, but there is a standard 3.5-mm audio jack to plug in headphones but USM mikes may work or not.

For video, the RCA jack sends video to any connected RCA video device.

4.2.6 GPIO Headers(Pins)

GPIO pins stands for general purpose of input output pins. These pins are used to connect any number of physical extensions with the Raspberry Pi. Raspberry Pi has pre-installed libraries that allow us to access the pins using programming languages like C, C++ or python.

4.2.7 Chips (Broadcom)

The most important component in a Raspberry Pi is chip that is Broadcom which is placed at the middle of the board. The chip consists of ARM11 processor running at 700 MHz and a Videocore4 GPU and can be over clocked to at least 800 MHz without a problem.

4.3 Raspberry Pi 4 Model B

The Raspberry Pi 4 uses a Broadcom BCM2711 SoC with a 1.5 GHz (later models: 1.8 GHz) 64-bit quad-core [ARM Cortex-A72](#) processor, with 1 MB shared L2 cache. Unlike previous models, which all used a custom [interrupt controller](#) poorly suited for virtualisation, the interrupt controller on this SoC is compatible with the ARM Generic Interrupt Controller (GIC) architecture 2.0, providing hardware support for interrupt distribution when using ARM virtualisation capabilities

4.4 Hardware Required for Raspberry Pi

Raspberry Pi can't start alone, it needs many others peripherals (hardware). There is brief description of the hardware requirements in the following section (Bates, 2014).

4.4.1 Power Supply

As mentioned already in above theory portion, Raspberry Pi needs 5V power supply. If supply exceeds 5V then it can't guaranteed to work properly. And the power supply also need to supply at least 500 milliamps (mA), and preferably more like 1 amp (A). If the supply is 500 mA or less, it is likely to have the malfunction of keyboard and mouse. It is not good idea to power the Raspberry Pi from USB port of computer and hub as they mostly provide current less than required. Hence, the Raspberry Pi requires a Micro-USB connection which is capable of supplying at least 700 mA (or 0.7 A) at 5V.

4.4.2 Storage

A separate hardware is required for the storage purpose in Raspberry Pi. For this, SD card is used, mostly 4 GB and 8 GB if needed. The operating system and all files are stored in the card. We can buy blank SD card and install operating system or buy a pre-installed one.

4.4.3 Input

External keyboard and mouse are required to provide input to the Raspberry Pi. No any additional software is needed to use the keyboard and mouse.

4.4.4 Monitor

We can use monitor or TV with HDMI port or DVI inputs as the screen for the Raspberry Pi. For DVI inputs, HDMI-to-DVI converters are required which can be finding easily in a market. Monitor is most important for the Raspberry Pi as it is the only way to see what we have done on it.

4.4.5 Network

As in laptop or computer, we can access to internet and network in Raspberry Pi as well. For that, we can use wired Ethernet connection which is easier option or Wi-Fi module to access Wi-Fi in the Raspberry Pi.

4.5 Power supply unit

A power supply (sometimes known as a power supply unit or PSU) is a device or system that supplies electrical or other types of energy to an output load or group of loads. The term is most commonly applied to electrical energy supplies, less often to mechanical ones, and rarely to others.

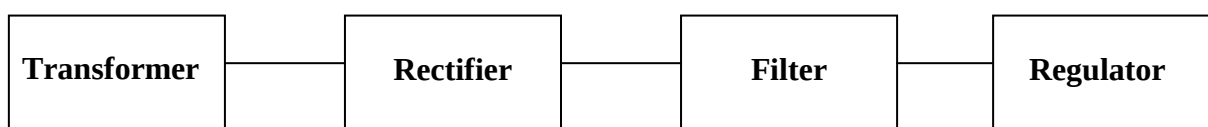


Fig4.2 Block diagram of power supply

The transformer steps up or steps down the input line voltage and isolates the power supply from the power line. The RECTIFIER section converts the alternating current input signal to a pulsating direct current. However, as you proceed in this chapter you will learn that pulsating dc is not desirable. For this reason a FILTER section is used to convert pulsating dc to a purer, more desirable form of dc voltage.

The final section, the REGULATOR, does just what the name implies. It maintains the output of the power supply at a constant level in spite of large changes in load current or input line voltages. Now that you know what each section does, let's trace an ac signal through the power supply. At this point you need to see how this signal is altered within each section of the power supply. Later on in the chapter you will see how these changes take place. An input signal of 115 volts ac is applied to the primary of the transformer.

The transformer is a step-up transformer with a turns ratio of 1:3. You can calculate the output for this transformer by multiplying the input voltage by the ratio of turns in the primary to the ratio of turns in the secondary; therefore, $115 \text{ volts ac} \times 3 = 345 \text{ volts ac}$ (peak-to-peak) at the output. Because each diode in the rectifier section conducts for 180 degrees of the 360-degree input, the output of the rectifier will be one-half, or approximately 173 volts of pulsating dc. The filter section, a network of resistors, capacitors, or inductors, controls the rise and fall time of the varying signal; consequently, the signal remains at a more constant dc level. You will see the filter process more clearly in the discussion of the actual filter circuits. The output of the filter is a signal of 110 volts dc, with ac ripple riding on the dc. The reason for the lower voltage (average voltage) will be explained later in this chapter. The regulator maintains its output at a constant 110-volt dc level, which is used by the electronic equipment (more commonly called the load).

4.5.1 Simple 5V power supply for digital circuits

- Brief description of operation: Gives out well-regulated +5V output, output current capability of 100 mA
- Circuit protection: Built-in overheating protection shuts down output when regulator IC gets too hot
- Circuit complexity: Very simple and easy to build
- Circuit performance: Very stable +5V output voltage, reliable operation
- Availability of components: Easy to get, uses only very common basic components
- Design testing: Based on datasheet example circuit, I have used this circuit successfully as part of many electronics projects
- Applications: Part of electronics devices, small laboratory power supply
- Power supply voltage: Unregulated DC 8-18V power supply
- Power supply current: Needed output current + 5 mA
- Component costs: Few dollars for the electronics components + the input transformer cost

4.6 WEBCAM:-

A **webcam** is a video camera which feeds its images in real time to a computer or computer network, often via USB, Ethernet or Wi-Fi. Their most popular use is the establishment of video links, permitting computers to act as videophones or videoconference stations. This common use as a video camera for the World Wide Web gave the webcam its name. Other popular uses include security surveillance and computer vision. Webcams are known for their low manufacturing cost and flexibility, making them the lowest cost form of video telephony. They have also become a source of security and privacy issues, as some built-in webcams can be remotely activated via spyware.

4.6.1 LENS

Webcams typically include a lens, an image sensor, support electronics, and may also include a microphone for sound. Various lenses are available, the most common in consumer-grade webcams being a plastic lens that can be screwed in and out to focus the camera. Fixed focus lenses, which have no provision for adjustment, are also available. As a camera system's depth of field is greater for small image formats and is greater for lenses with a large f-number (small aperture), the systems used in webcams have a sufficiently large depth of field that the use of a fixed focus lens does not impact image sharpness to a great extent.

Image sensors can be CMOS or CCD, the former being dominant for low-cost cameras, but CCD cameras do not necessarily outperform CMOS-based cameras in the low cost price range. Most consumer webcams are capable of providing VGA resolution video at a frame rate of 30 frames per second. Many newer devices can produce video in multi-megapixel resolutions, and a few can run at high frame rates such as the PlayStation Eye, which can produce 320×240 video at 120 frames per second.

Support electronics read the image from the sensor and transmit it to the host computer. The camera pictured to the right, for example, uses a Sonix SN9C101 to transmit its image over USB. Typically, each frame is transmitted uncompressed in RGB or YUV or compressed as JPEG. Some cameras, such as mobile phone cameras, use a CMOS sensor with supporting electronics "on die", i.e. the sensor and the support electronics are built on a single silicochip to save space and manufacturing costs.

4.6.2 PORTABILITY:

Most webcams feature built-in microphones to make video calling and videoconferencing more convenient. The USB video device class (UVC) specification allows for interconnectivity of webcams to computers without the need for proprietary device drivers. Microsoft Windows XP SP2, Linux and Mac OS X (since October 2005) have UVC support built in and do not require extra device drivers, although they are often installed to add additional features.

4.6.3 FEATURES (LOGITECH WEBCAM C100)

- Plug-and-play setup (UVC)
- Video capture: Up to 640 x 480 pixels
- Photos: Up to 1.3 megapixels (software enhanced)
- Frame rate: Up to 30 frames per second (with recommended system)
- Hi-Speed USB 2.0 certified
- Fixed focus
- Universal clip fits notebooks, LCD or CRT monitor



Webcam

4.7 Image Processing on Raspberry Pi

Today image processing are used in various techniques, this paper presents the implementation of image processing operations on Raspberry Pi. The Raspberry Pi is a basic embedded system and being a low cost a single board computer used to reduce the complexity of systems in real time applications. This platform is mainly based on python. Raspberry pi consist of Camera slot Interface (CSI) to interface the raspberry pi camera. Here, the Dark and Low contrast images captured by using the Raspberry Pi camera module are enhanced in order to identify the particular region of image. This concept is used in the real time application of MAV, The MAVs are basically used to capture images and videos through the Raspberry pi camera module. Because of its credit card sized (small) and less weight in the design. However, the image captured by MAVs will consist of unwanted things due to atmospheric conditions; hence it is necessary to remove noise present in the MAVs images.

4.8 INTRODUCTION

The image processing is a form of signal processing where the input is an image, photograph or video frame, the output of an image processing may be either an image or a video frame or a set of characteristics or parameters related to the image. The acquisition of digital image usually suffers from undesirable camera shakes and due to unstable random camera motions. Hence image enhancement algorithms are required to remove these unwanted camera shakes. This image processing concepts are implemented in Raspberry pi in the application of MAV. The Raspberry Pi is a basic embedded system having a credit card-sized single board computers developed in the UK by the Raspberry Pi Foundation. The Raspberry Pi is based on the Broadcom BCM2835 system on a chip (SOC) which includes an ARM1176JZF-S Core (ARM V6K)700 MHz CPU processor, Broadcom Video Core IV GPU having 17 pins, 3.5W of power, and 512 MB of RAM memory. The Raspberry Pi system has Secure SD card reader (models A and B) or Micro SD card reader (models A+ and B+) sockets for boot media and persistent storage. The system provides Debian Linux operating system Raspbian image for download. Python is used as main programming language for raspberry pi. A micro air vehicle (MAV) is a remote-controlled, Unmanned Aircraft Vehicle

(UAV) significantly smaller than typical UAVs that have a size restriction. UAV is an aircraft without a human pilot. Its flight is controlled either autonomously on board computers or by the remote control of a pilot on the ground or in another vehicle. By having a Raspberry Pi camera module available on a MAV the efficiency of this air vehicle increases and new fields of applications are available. It is needed in military Operations, in which targets have to be identified. Such identification is often done by a human on ground, to reduce the probability of mistakes. But a Raspberry Pi camera module is also helpful if a MAV shall autonomously fly through an arch.

4.9 BASIC CONCEPT OF IMAGE PROCESSING

In general, any digital image processing algorithm consists of three stages: input, processor and output. In the input stage image is captured by a camera. It sent to a particular system to focus on a pixel of image that's gives, its output as a processed image.

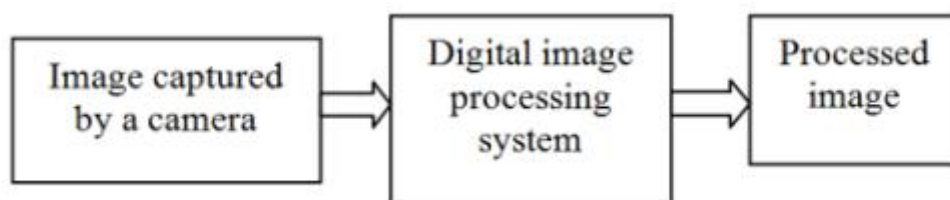


Fig4.3

General Block diagram of image processing

4.10 SYSTEM HARDWARE DESIGN

The Raspberry Pi board is the central module of the whole embedded image capturing and processing system as given in fig. 2. Its main parts include: main processing chip unit, memory, power supply HDMI Out i.e VGA display, Ethernet port, and USB ports.

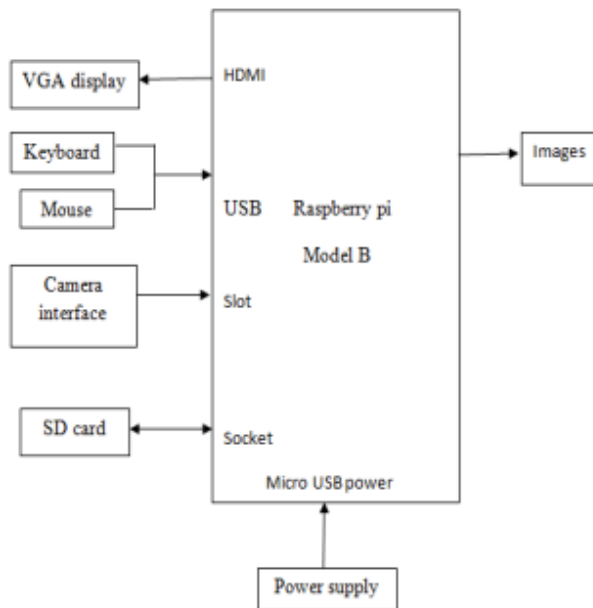


Fig4.4 System Block Diagram

4.11 Ultrasonic sensing

Ultrasonic sensing is one of the best ways to sense proximity and detect levels with high reliability.

Our technical support gets emails all of the time about how our sensors work and what environments our sensors work (or don't work) in.

This guide was created as an introduction to ultrasonic sensing, it's principles, and how ultrasonic sensors work in your applications.

At a top level, you need to know what an ultrasonic sensor is...

4.12 What is an ultrasonic sensor?

An ultrasonic sensor is an instrument that measures the distance to an object using ultrasonic sound waves.

An ultrasonic sensor uses a transducer to send and receive ultrasonic pulses that relay back information about an object's proximity.

High-frequency sound waves reflect from boundaries to produce distinct echo patterns.

4.13 How Ultrasonic Sensors Work

Ultrasonic sensors work by sending out a sound wave at a frequency above the range of human hearing. The transducer of the sensor acts as a microphone to receive and send the ultrasonic sound. Our ultrasonic sensors, like many others, use a single transducer to send a pulse and to receive the echo. The sensor determines the distance to a target by measuring time lapses between the sending and receiving of the ultrasonic pulse.

The working principle of this module is simple. It sends an ultrasonic pulse out at 40kHz which travels through the air and if there is an obstacle or object, it will bounce back to the sensor. By calculating the travel time and the speed of sound, the distance can be calculated.

Ultrasonic sensors are a great solution for the detection of clear objects. For liquid level measurement, applications that use infrared sensors, for instance, struggle with this particular use case because of target translucence.

For presence detection, ultrasonic sensors detect objects regardless of the color, surface, or material (unless the material is very soft like wool, as it would absorb sound.)

To detect transparent and other items where optical technologies may fail, ultrasonic sensors are a reliable choice.

4.14 Using Multiple Sensors & Avoiding Disruption

When using multiple sensors in an application, it's important to connect them in a way that will allow you to avoid issues like crosstalk or any other interference.

To prevent the disruption of the ultrasonic signals coming from your sensor, it's important to keep the face of the ultrasonic transducer clear of any obstructions.

Common obstructions include:

- Dirt
- Snow
- Ice

Other Condensation

For this particular use case, we offer our Self Cleaning sensors.

They are intended specifically for applications requiring the resistance of condensation in high moisture environments, our self-cleaning function is designed to run continuously in order for the self-cleaning feature to be active.

4.15 How are Ultrasonic Sensors used?

Our ultrasonic distance, level, and proximity sensors are commonly used with microcontroller platforms like Raspberry Pi, ARM, PIC, Arduino, Beagle Board, and more.

Ultrasonic sensors transmit sound waves toward a target and will determine its distance by measuring the time it took for the reflected waves to return to the receiver.

This sensor is an electronic device that will measure the distance of a target by transmitting ultrasonic sound waves, and then will convert the reflected sound into an electrical signal.

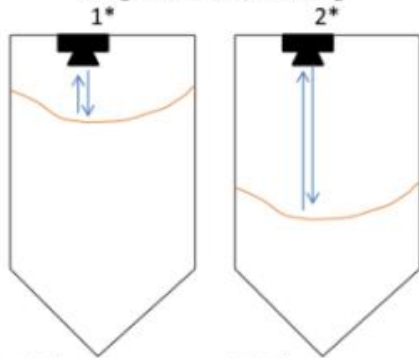
Our sensors are often used as proximity sensors.

Ultrasonic sensors are also used in obstacle avoidance systems, as well as in manufacturing.

Our ShortRange sensors offer the opportunity for closer range detection where you may need a sensor that ranges objects as close to 2cm. These are also built with very low power requirements in mind, as well as environments where noise rejection is necessary.

Grain Detection

Diagrams 1 & 2, Working



1. 1/2 Full-
There is typically a level enough surface for sound to be reflected back to the sensor, thus yielding the distance to the grain being reported.

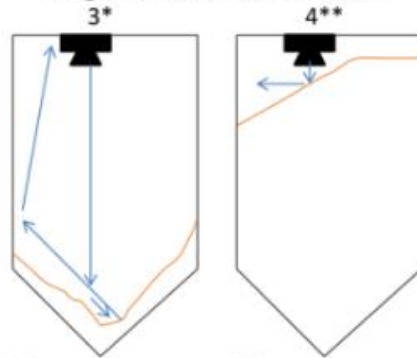
2. 1/2 Full-
There is typically a level enough surface for sound to be reflected back to the sensor, thus yielding the distance to the grain being reported.

*For Best Results

For best results with grain, use the MB7954 Snap-On horn available at this link http://www.maxbotix.com/Ultrasonic_Sensors/MB7974.htm. This product increased the sensitivity of the sensors which increases the reliability of the detection to grain, a soft target. This product accessory doesn't guarantee 100% accurate results; however, it has been proven to increase the reliability of the detection of grain.



Diagrams 3 & 4, Potential Issues



3. Near empty-
Shape of the grain may be deflected away from the sensor because it is not a level surface. This may cause a max range reading to be reported.

4. Full-
Shape of the grain up close may cause the sound to be deflected away from the sensor. This may cause a max range reading to be reported.

**Upclose Detection

Applications requiring 100% reading-to-reading reliability should not use MaxSonar sensors at a distance closer than the minimum reported distance. Although most users have found MaxSonar sensors to work reliably from 0mm to the minimum reported distance in many applications, MaxBotix Inc., does not guarantee operational reliability for objects closer than the minimum reported distance. Because of physics relating to ultrasonics, these sensors are unable to achieve 100% reliability at close distance.

Recommended Sensor Mounting-

- Typically the best mounting is to have the sensor offset from the center of the bin by at least several inches.
- Smooth walls are important because cube corner reflections can cause incorrect range readings. You can read more about cube corner reflections at this link http://en.wikipedia.org/wiki/Corner_reflector.
- Do not mount the sensor too close to the side of the bin (typically closer than 18 inches even when the bin has smooth walls) because phase cancelation of the sound wave may cause incorrect range readings to occur.

CHAPTER 5

SOFTWARE DESCRIPTION

5.1 Start up with Raspberry Pi

5.1.1 Installation of operating system on Raspberry Pi

Raspberry Pi is a small computer; hence operating system (OS) should be installed. As the Raspberry doesn't have hard drive, OS is installed in the external memory. For that, memory card (SD card) is used for the installation of operating system and all the required software and supporting files are stored in the same SD card.

There are different types of operating system but we preferred to talk about NOOBS (New Out of the Box Software) as it is suitable for the beginners. We can either buy a preinstalled SD card or empty SD card. In pre-installed SD card, NOOBS is already copied and ready to boot. The detail description of installation of operating system in a blank SD card in step wise is given in Appendix B.

5.1.2 Use of laptop screen, keyboard and mouse for Raspberry Pi

If you don't have suitable screen for the Raspberry Pi, we can use the monitor of a laptop. Besides, we can use keyboard and mouse of laptop as well for Raspberry Pi. For that, two software are needed to install in a laptop which screen is going to use for Raspberry Pi. Installation and configuration of the software is discussed briefly in the following sections.

Xming

This is the first software need to be installed which can be downloaded from the link, [Download_Xming](#) and install it in the laptop. After completion of installation, run the application called 'XLaunch' and verify that the settings are as shown in three following figures:

Finally, click "Next" button which is shown below it goes to last configuration dialogue box where "Finish" button is needed to click for the completion of the setting. After Completion of configuration, double click the application named as "Xming".

1 5.2 PuTTY

This is the primary software need to be installed. It can be downloaded in the provided link as follow, [Download Putty](#). As, it is downloaded, it needs to be installed following some few normal steps of installation. For Configuration, double click the icon of Putty after the completion of installation and enter the IP address of Raspberry Pi as shown.

Finally, click on "open" button and as a result; window will be displayed as shown.

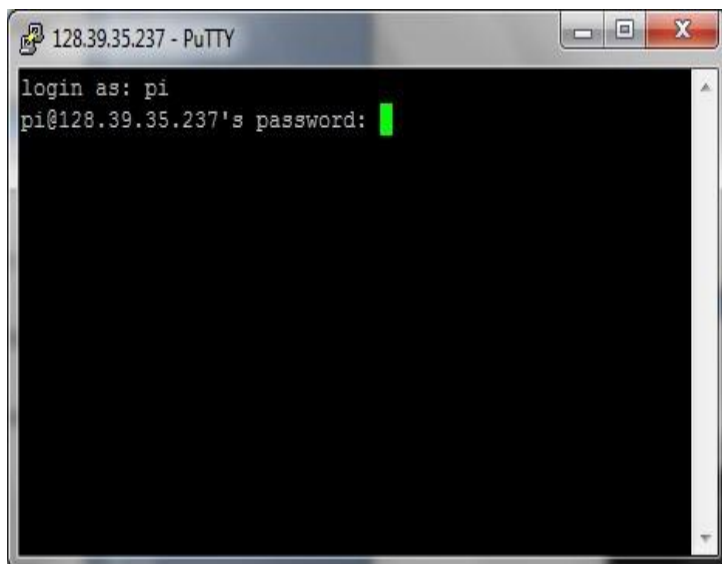


Figure 5.1: PuTTY login windows

In this window, username and password of the Raspberry Pi should be entered. The default username for a Raspberry Pi is pi. Press enter after entering correct username and password and you will see the windows where you should enter the text "lesson" as shown in Figure 12-7.

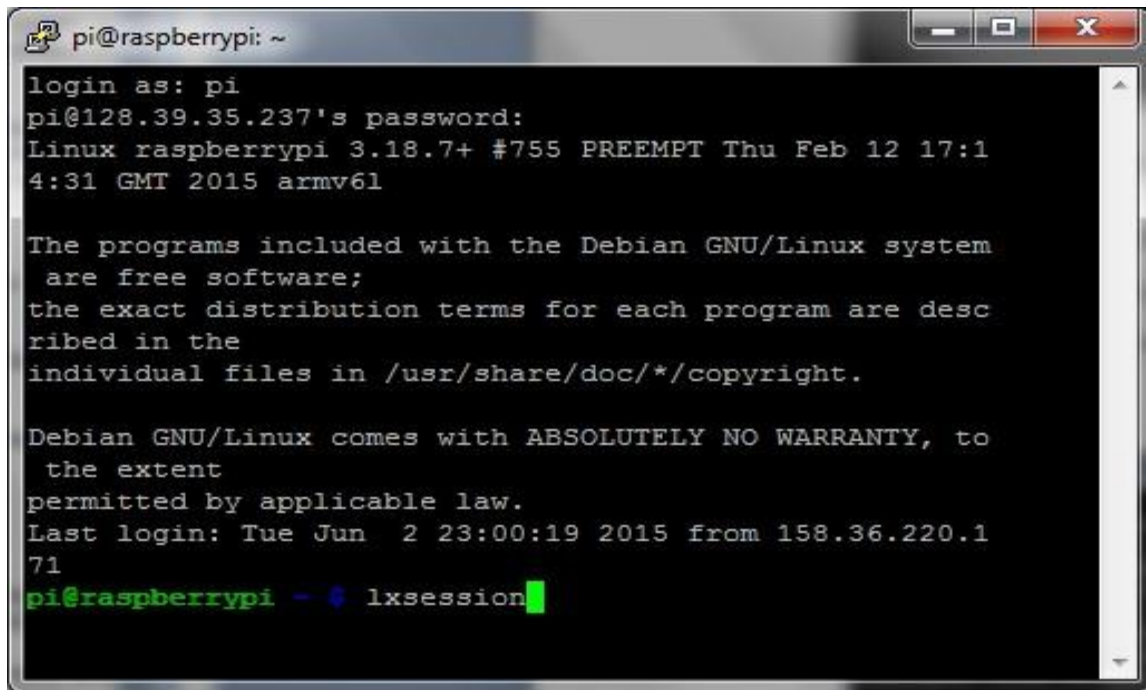


Figure 5.2: PuTTY success windows

Then, you will be able to see the desktop of Raspberry Pi in the laptop as shown in Figure 12-8 and mouse and keyboard can be used for the Raspberry Pi.

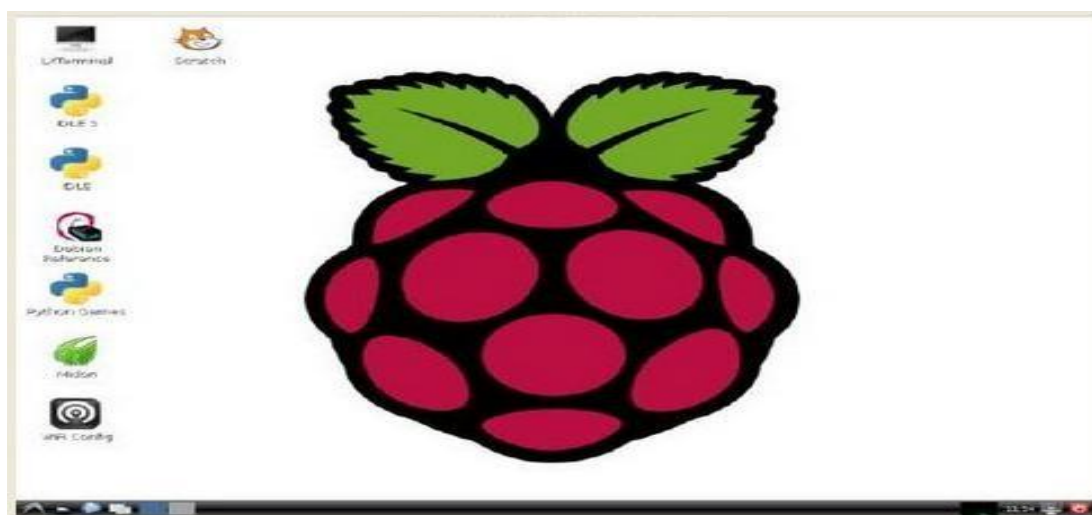


Figure 5.3: Raspberry Pi Desktop on your Laptop Display

5.3 Installation of required applications on Raspberry Pi

There are many applications that are needed to install in the Raspberry Pi for the completion of the thesis. For data logging, MySQL apache5 and php my admin are needed to install whereas for the web-page development, PHP is needed to install. Web page is used for the monitoring and managing purpose

5.3.1 Raspbian OS

Raspberry Pi OS (formerly Raspbian) is a Debian-based operating system for Raspberry Pi. Since 2015, it has been officially provided by the Raspberry Pi Foundation as the primary operating system for the Raspberry Pi family of compact single-board computers. The first version of Raspbian was created by Mike Thompson and Peter Green as an independent project The initial released build was completed on July 15, 2012'

Raspberry Pi OS is highly optimized for the Raspberry Pi line of compact single-board computers with ARM CPUs. It runs on every Raspberry Pi except the Pico microcontroller. Raspberry Pi OS uses a modified LXDE as its desktop environment with the Openbox stacking window manager, along with a unique theme. The distribution is shipped with a copy of the algebra program Wolfram Mathematica and a version of Minecraft called Minecraft: Pi Edition (note that Minecraft: Pi Edition is no longer installed as of the Debian bullseye update) as well as a lightweight version of the Chromium web browser.

5.3.2 User Interface

Raspberry Pi OS's desktop environment, PIXEL, looks similar to many common desktops, such as macOS, Microsoft Windows, and is based on LXDE. The menu bar is positioned at the top and contains an application menu and shortcuts to Terminal, Chromium, and File Manager. On the right is a Bluetooth menu, a Wi-Fi menu, volume control, and a digital clock.

5.3.3 Components

PCManFM is a file browser allowing quick access to all areas of the computer, and was redesigned in the first Raspberry Pi OS Buster release (2019-06-20).

Raspberry Pi OS originally used Epiphany as the web browser, but switched to Chromium with the launch of its redesigned desktop.

Raspberry Pi OS comes with many beginner IDEs, such as Thonny Python IDE, Mu Editor, and Greenfoot. It also ships with educational software like Scratch and Bookshelf.

5.3.4 Version Size

The Raspberry Pi documentation recommends at least a 4GB microSD card for Raspberry Pi OS Lite, an 8GB microSD card for Raspberry Pi OS, and a 16GB microSD card for Raspberry Pi OS Full.[19] The image files themselves are 1,875MB, 3,980MB, and 8,603MB respectively.

5.4 Python

5.4.1 What is Python?

Python: Dynamic programming language which supports several different programming paradigms:

- Procedural programming
- Object oriented programming
- Functional programming

Standard: Python byte code is executed in the Python interpreter (similar to Java) → platform independent code

5.4.2 Why Python?

Extremely versatile language Website development, data analysis, server maintenance, numerical analysis,

- Syntax is clear, easy to read and learn (almost pseudo code)
- Common language
- Intuitive object oriented programming

- Full modularity, hierarchical packages
- Comprehensive standard library for many tasks
- Big community Simply extendable via C/C++, wrapping of C/C++ libraries
- Focus: Programming speed

5.4.3 History

Start implementation in December 1989 by Guido van Rossum (CWI)

- 16.10.2000: Python 2.0
- Unicode support
- Garbage collector
- Development process more community oriented
- 3.12.2008: Python 3.0
- Not 100% backwards compatible 2007 & 2010 most popular programming language (TIOBE Index)
- Recommendation for scientific programming (Nature News, NPG, 2015)
- Current version: Python 3.9.2
- Python2 is out of support

5.4.4 Is Python fast enough?

- For user programs: Python is fast enough!
- Most parts of Python are written in C
- For compute intensive algorithms: Fortran, C, C++ might be better
- Performance-critical parts can be re-implemented in C/C++ if necessary
- First analyses, then optimize!

5.4.5 Conclusion

Python is a fully-fledged programming language. Any code you download from github, the Python Package Index, or anywhere else, can be malicious and would almost certainly go unnoticed by most people's firewalls.

5.5 OpenCV

OpenCV is a cross-platform library using which we can develop real-time computer vision applications. It mainly focuses on image processing, video capture and analysis including features like face detection and object detection. Let's start the chapter by defining the term "Computer Vision"

5.5.1 Computer Vision

Computer Vision can be defined as a discipline that explains how to reconstruct, interrupt, and understand a 3D scene from its 2D images, in terms of the properties of the structure present in the scene. It deals with modeling and replicating human vision using computer software and hardware.

Computer Vision overlaps significantly with the following fields:

- Image Processing: It focuses on image manipulation.
- Pattern Recognition: It explains various techniques to classify patterns.
- Photogrammetry: It is concerned with obtaining accurate measurements from images.

5.5.2 Computer Vision Vs Image Processing

Image processing deals with image-to-image transformation. The input and output of image processing are both images.

Computer vision is the construction of explicit, meaningful descriptions of physical objects from their image. The output of computer vision is a description or an interpretation of structures in 3D scene.

5.5.3 Applications of Computer Vision

Here we have listed down some of major domains where Computer Vision is heavily used. **Robotics Application**

- Localization — Determine robot location automatically
- Navigation
- Obstacles avoidance
- Assembly (peg-in-hole, welding, painting)

- Manipulation (e.g. PUMA robot manipulator)
- Human Robot Interaction (HRI): Intelligent robotics to interact with and serve people

5.5.4 Medicine Application

Classification and detection (e.g. lesion or cells classification and tumor detection)

- 2D/3D segmentation
- 3D human organ reconstruction (MRI or ultrasound)
- Vision-guided robotics surgery

5.5.5 Industrial Automation Application

- Industrial inspection (defect detection)
- Assembly
- Barcode and package label reading
- Object sorting
- Document understanding (e.g. OCR)

5.5.6 Security Application

- Biometrics (iris, finger print, face recognition)
- Surveillance — Detecting certain suspicious activities or behaviors Transportation

5.5.7 Application

- Autonomous vehicle
- Safety, e.g., driver vigilance monitoring

5.5.8 Features of OpenCV Library

Using OpenCV library, you can –

- Read and write images
- Capture and save videos
- Process images (filter, transform)
- Perform feature detection

- Detect specific objects such as faces, eyes, cars, in the videos or images.
- Analyze the video, i.e., estimate the motion in it, subtract the background, and track objects in it. OpenCV was originally developed in C++. In addition to it, Python and Java bindings were provided. OpenCV runs on various Operating Systems such as windows, Linux, OSX, FreeBSD, Net BSD, Open BSD, etc.

This tutorial explains the concepts of OpenCV with examples using Java bindings.

5.5.9 OpenCV Library Modules

Following are the main library modules of the OpenCV library.

Core Functionality

This module covers the basic data structures such as Scalar, Point, Range, etc., that are used to build OpenCV applications. In addition to these, it also includes the multidimensional array Mat, which is used to store the images. In the Java library of OpenCV, this module is included as a package with the name `org.opencv.core`.

5.5.10 Image Processing

This module covers various image processing operations such as image filtering, geometrical image transformations, color space conversion, histograms, etc. In the Java library of OpenCV, this module is included as a package with the name `org.opencv.imgproc`.

5.5.11 Video

This module covers the video analysis concepts such as motion estimation, background subtraction, and object tracking. In the Java library of OpenCV, this module is included as a package with the name `org.opencv.video`.

5.5.12 Video I/O

This module explains the video capturing and video codecs using OpenCV library. In the Java library of OpenCV, this module is included as a package with the name `org.opencv.videoio`.

5.5.13 calib3d

This module includes algorithms regarding basic multiple-view geometry algorithms, single and stereo camera calibration, object pose estimation, stereo correspondence and elements of 3D reconstruction. In the Java library of OpenCV, this module is included as a package with the name `org.opencv.calib3d`.

5.5.14 features2d

This module includes the concepts of feature detection and description. In the Java library of OpenCV, this module is included as a package with the name `org.opencv.features2d`.

5.5.15 Objdetect

This module includes the detection of objects and instances of the predefined classes such as faces, eyes, mugs, people, cars, etc. In the Java library of OpenCV, this module is included as a package with the name `org.opencv.objdetect`.

5.5.16 Highgui

This is an easy-to-use interface with simple UI capabilities. In the Java library of OpenCV, the features of this module is included in two different packages namely, `org.opencv.imgcodecs` and `org.opencv.videoio`.

5.5.17 OpenCV — Storing Images

To capture an image, we use devices like cameras and scanners. These devices record numerical values of the image (Ex: pixel values). OpenCV is a library which processes the digital images, therefore we need to store these images for processing. The `Mat` class of OpenCV library is used to store the values of an image. It represents an n-dimensional array and is used to store image data of grayscale or color images, voxel volumes, vector fields, point clouds, tensors, histograms, etc. This class comprises of two data parts: the header and a pointer

- Header: Contains information like size, method used for storing, and the address of the matrix (constant in size).
- Pointer: Stores the pixel values of the image (Keeps on varying).

5.5.18 the Mat Class

The OpenCV Java library provides this class with the same name (Mat) within the package org.opencv.core.

5.5.19 Constructors

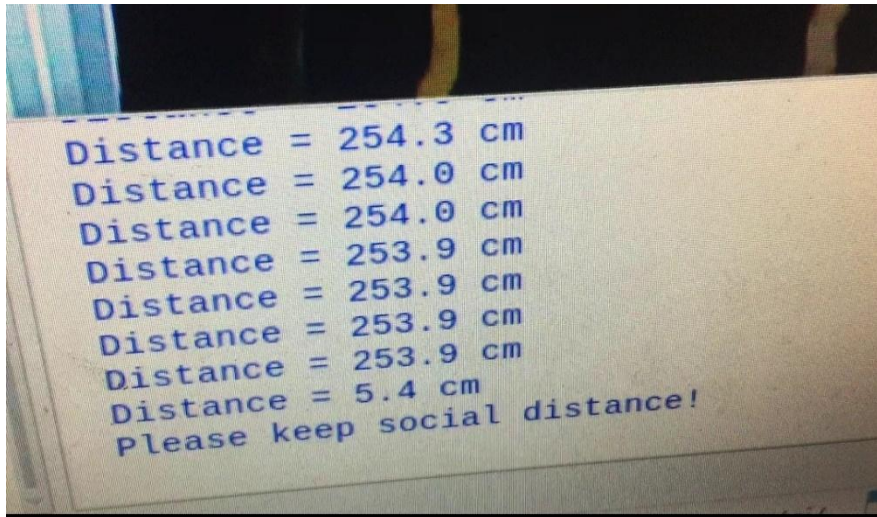
The Mat class of OpenCV Java library has various constructors, using which you can construct the Mat object.

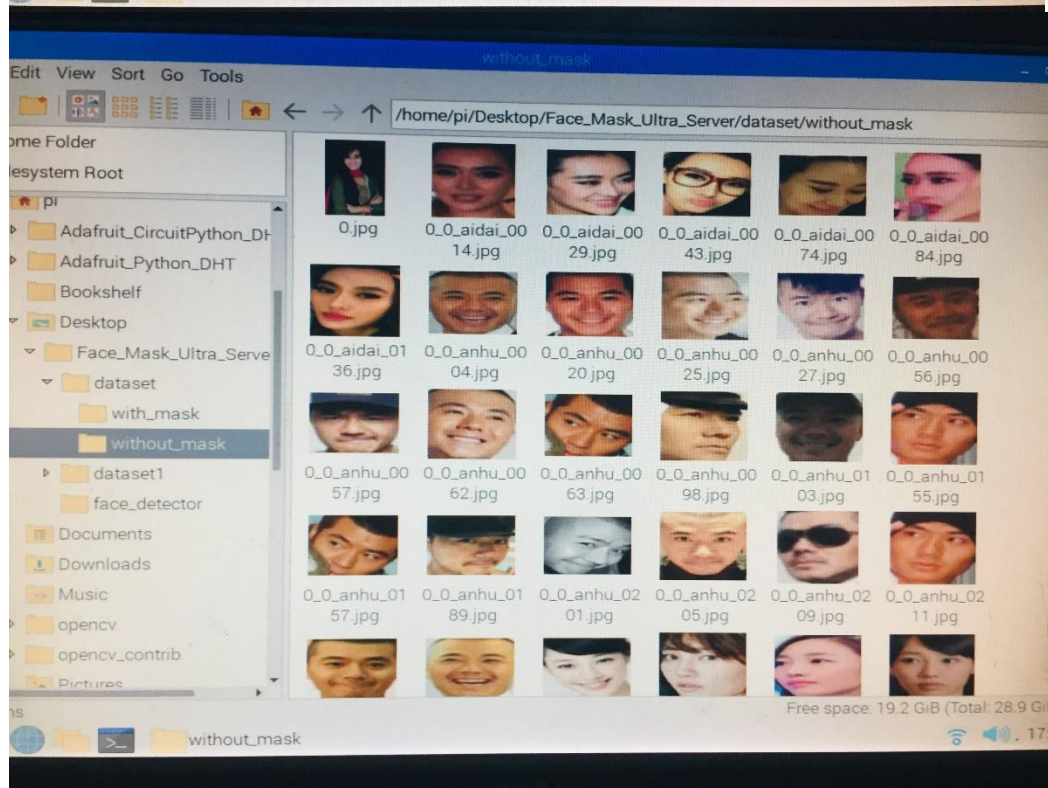
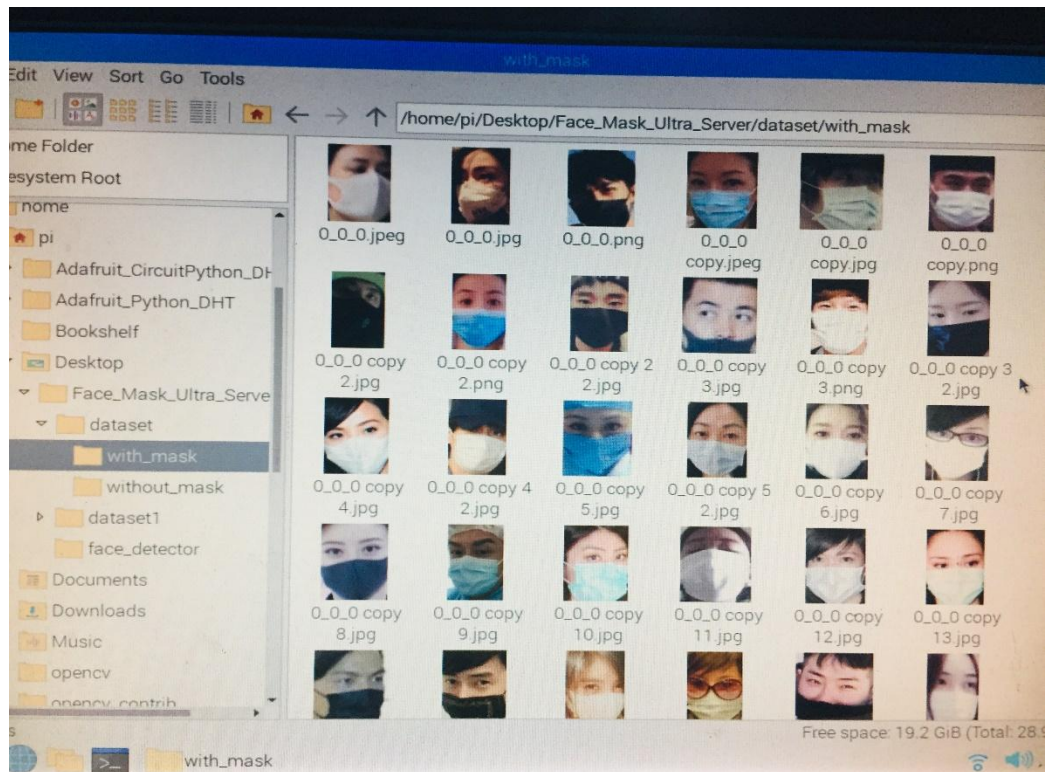
1. This is the default constructor with no parameters in most cases. We use this to constructor to create an empty matrix and pass this to other OpenCV methods.
 2. `Mat(int rows, int cols, int type)` This constructor accepts three parameters of integer type representing the number of rows and columns in a 2D array and the type of the array (that is to be used to store data).
 3. `Mat(int rows, int cols, int type, Scalar s)` Including the parameters of the previous one, this constructor additionally accepts an object of the class Scalar as parameter.
 4. `Mat(Size size, int type)`
3. OpenCV —
Storing Imag

CHAPTER 6

RESULTS AND CONCLUSION

6.1 RESULT:





6.2 CONCLUSIONS:

In this analysis, we got described an epitome system to scan written communication and hand-held objects for serving to the blind people. To extract text regions from advanced backgrounds, we have got projected a totally distinctive text localization formula supported models of stroke orientation, and edge distributions. The corresponding feature maps estimate the worldwide structural feature of text at every element. Block patterns project the projected feature maps of an image patch into a feature vector. A adjacent character grouping is performed to calculate candidates of text patches prepared for text classification. Associate Ada boost learning model is used to localize text in camera-based footage. OCR is utilized to perform word recognition on the localized text regions and work into audio output for blind users. Throughout this analysis, the camera acts as input for the paper. As a result of the Raspberry Pi board is high-powered the camera starts streaming. The streaming data square measure attending to be displayed on the screen victimization interface application. Once the item for text reading is placed earlier than the camera then the capture button is clicked to supply image to the board. Mistreatment Tesseract library the image square measure attending to be converted into data, and conjointly the data detected from the image square measure attending to be shown on the standing bar. The obtained data square measure attending to be pronounced through.

REFERENCES:

- [1] X. Chen and A. L. Yuille, "Detecting and reading text in natural scenes," in Proc. Comput. Vision Pattern Rec-ognit., 2004, vol. 2, pp. II-366–II-373.
- [2] S. Kumar, R. Gupta, N. Khanna, S. Chaudhury, and S. D. Joshi, "Text extraction and document image seg-mentation using matched wavelets and MRF model," IEEE Trans Image Process., vol. 16, no. 8, pp. 2117– 2128, Aug. 2007.
- [3] K. Kim, K. Jung, and J. Kim, "Texture-based ap-proach for text detection in images using support vec-tor machines and continuously adaptive mean shift algorithm," IEEE Trans. Pattern Anal. Mach. Intell., vol. 25, no. 12, pp. 1631–1639, Dec. 2003.
- [4] N. Giudice and G. Legge, "Blind navigation and the role of technology," in The Engineering Handbook of Smart Technology for Aging, Disability, and Indepen-dence, A. A. Helal, M. Mokhtari, and B. Abdulrazak, Eds. Hoboken, NJ, USA: Wiley, 2008.
- [5] World Health Organization. (2009). 10 facts about blindness and visual impairment.
- [6] Advance Data Reports from the National Health In-terview Survey (2008).
- [7] International Workshop on Camera-Based Docu-ment Analysis and Recognition (CBDAR 2005, 2007, 2009, 2011).
- [8] X. Chen, J. Yang, J. Zhang, and A. Waibel, "Auto-matic detection and recognition of signs from natural scenes," IEEE Trans. Image Process., vol. 13, no. 1, pp. 87–99, Jan. 2004.
- [9] D. Dakopoulos and N. G. Bourbakis, "Wearable ob-stacle avoidance electronic travel aids for blind: A sur-vey," IEEE Trans. Syst., Man, Cybern., vol. 40, no. 1, pp. 25–35, Jan. 2010.

- [10] B. Epshtein, E. Ofek, and Y. Wexler, "Detecting text in natural scenes with stroke width transform," in Proc. Comput. Vision Pattern Recognit., 2010, pp. 2963–2970.
- [11] Y. Freund and R. Schapire, "Experiments with a new boosting algorithm," in Proc. Int. Conf. Machine Learning, 1996, pp. 148–156.
- [13] An overview of the Tesseract OCR (optical character recognition) engine, and its possible enhancement for use in Wales in a pre-competitive research stage Prepared by the Language Technologies Unit (Canol-fan Bedwyr), Bangor University April 2008.
- [14] A. Shahab, F. Shafait, and A. Dengel, "ICDAR 2011 robust reading competition: ICDAR Robust Reading Competition Challenge 2: Reading text in scene images," in Proc. Int. Conf. Document Anal. Recognit., 2011, pp. 1491–1496.
- [15] KReader Mobile User Guide, knfb Reading Technology Inc. (2008).
- [16] S. M. Lucas, "ICDAR 2005 text locating competition results," in Proc. Int. Conf. Document Anal. Recognit., 2005, vol. 1, pp. 80–84.