VISVESVARAYA TECHNOLOGICAL UNIVERSITY

BELAGAVI-590018, KARNATAKA.

A PROJECT REPORT

On

"FACE RECOGNITION BASED VOICE ASSISTANT"

Submitted in Partial Fulfillment for the Award of the Degree

of

BACHELOR OF ENGINEERING

IN

COMPUTER SCIENCE & ENGINEERING

Submitted By:

1SG17CS003	**ADNAN ANJUM**
1SG17CS008	**AMAN PANDEY**

Under the Guidance of

Prof. Nanda M B
Assistant Professor, CSE, SCE

Department of Computer Science and Engineering
SAPTHAGIRI COLLEGE OF ENGINEERING
14/5, Chikkasandra, Hesarghatta Main Road
Bengaluru – 560057.

2020-2021

SAPTHAGIRI COLLEGE OF ENGINEERING

(Affiliated to VTU, Belagavi, Approved bu AICTE, New Delhi)
ISO 9001:2015 & ISO 14001:2015 Certified Institute
Accredited By NAAC with "A" Grade
#14/5, Chikkasandra, Hesaraghatta Main Road, Bangalore – 560057

Department of Computer Science and Engineering

Certificate

Certified that the Project Work entitled **"Face Recognition based Voice Assistant"** carried out by **ADNAN ANJUM (1SG17CS003), AMAN PANDEY (1SG17CS008),** bonafide students of **Sapthagiri College of Engineering**, in partial fulfillment for the award of **Bachelor of Engineering** in **Computer Science and Engineering** of **Visvesvaraya Technological University**, **Belagavi** during the academic year 2020-2021. It is certified that all corrections/suggestions indicated for Internal Assessment have been incorporated in the report deposited in the Department Library. The project report has been approved as it satisfies the academic requirements in respect of **Project Work Phase II (17CSP85)** prescribed for the said degree.

Signature of the Guide	Signature of the HOD	Signature of the Principal
Prof. Nanda M B	**Dr. Kamalakshi Naganna**	**Dr. H Ramakrishna**
Assistant Professor	**Professor & Head**	**Principal**

EXTERNAL EXAMINATION:

Name of the Examiners	Signature with Date
1._____	_____
2._____	_____

ACKNOWLEDGEMENT

The satisfaction and euphoria that accompany the completion of any task would be incomplete without the mention of the people who made it possible, whose constant guidance and encouragement ground my efforts with success.

I consider it is a privilege to express my gratitude and respect to all those who guided me in completion of technical seminar.

I am, grateful to thank our Principal **Dr. H Ramakrishna**, Sapthagiri College of Engineering, who patronized throughout our career & for the facilities provided to carry out this work successfully.

It's a great privilege to place on record my deep sense of gratitude to our beloved HOD **Dr. Kamalakshi Naganna** of Computer Science & Engineering, who patronized throughout our career & for the facilities provided to carry out this work successfully.

I am also grateful to thank my Technical seminar Guide Prof. **Nanda M B, Assistant Professor, CSE, SCE** of CSE department for **his/her** invaluable support and guidance.

I would also like to thank the teaching and non-teaching staff members who have helped me directly or indirectly during the technical seminar.

Lastly but most importantly I would like thank my family and friends for their co-operation and motivation to complete this seminar successfully.

ADNAN ANJUM **1SG17CS003**

AMAN PANDEY **1SG17CS008**

ABSTRACT

Evolution of Communication and Technology are quite old and is constantly growing and changing. A smart environment is an environment which can identify people, interpret their actions and react accordingly. A face recognition system is one of the biometric information processes, its applicability is easier and working range is larger than others. The attempt has been made to develop an "Intelligent Personal Voice Assistant using Python" which helps people to do some basic tasks with their voice(speech), extract information and perform tasks on their desktop. Along with the voice assistant, a "Facial Recognition System" has been added to verify the identity of the voice assistant user. The program takes a face image and measure the characteristics such as the distance between the eyes, the length of nose etc. With growing the use of data and almost being completely dependent on data, security is needed more than before and what could be better than a biometric security feature.

SAPTHAGIRI COLLEGE OF ENGINEERING

(Affiliated to VTU, Belagavi, Approved bu AICTE, New Delhi)
ISO 9001:2015 & ISO 14001:2015 Certified Institute
Accredited By NAAC with "A" Grade
#14/5, Chikkasandra, Hesaraghatta Main Road, Bangalore – 560057

Department of Computer Science and Engineering

DECLARATION

We **ADNAN ANJUM(1SG17CS003)** and **AMAN PANDEY(1SG17CS003) ,** bonafide students of **Sapthagiri College of Engineering,** hereby declare that the project entitled "**FACE RECOGNITION BASED VOICE ASSISTANT** " submitted in partial fulfilment for the award of Bachelor of Engineering in **Computer Science & Engineering** of the Visvesvaraya Technological University, Belgaum during the year 2020-2021 is our original work and the project has not formed the basis for the award of any other degree, fellowship or any other similar titles.

Name & Signature of the Student with date

1)
2)
3)
4)

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

Chapter 1

INTRODUCTION

1.1. Background

As there has been a rapid increase in the uses of voice assistant and almost everyone currently uses a voice assistant. A voice assistant can do any tasks by just commanding it in the normal natural language. Nowadays a single voice assistant are available for the whole family or the persons living under a single roof. A problem did arise of getting a lower privacy. For any person privacy is much necessary and the person actually owning the assistant can give away its personal information if someone from a family or friend, just asks a simple question about the day's plan or commanding it to send someone a mail or a message could be a problem to the owner, as it was done without the permission. So, to remove this flaw of the current voice assistants, the implementation is done keeping data security of the user in mind and still using the full functionality of voice assistants. A facial recognition, as an additional biometric feature, was introduced in the implementation of the voice assistant to remove the above-mentioned issue.

1.2. Overview

Face Recognition is the task of recognizing a person based on its facial image or a live camera feed. It has acquired its popularity in past two decades. Facial Recognition Systems are built on computer programs that detects and analyses human face for the purpose of identifying and authorizing them accordingly. The program takes a face image and measure the characteristics such as the distance between the eyes, the length of nose etc. With growing the use of data and almost being completely dependent on data, security is needed more than before and what could be better than a biometric security feature. We can implement this biometric feature using several machine learning algorithms and deep learning techniques and hence provide the data from the assistant to the relevant user.

1.3. Objective

With all the current voice assistants available such as Alexa, Google assistant and various others, there is a little security flaw. As the current voice assistants, all being voice controlled, none of them identifies the user who commands it. This issue of not identifying the user may rise to a problem of giving away the data to the user who shouldn't have it. For example, a

command “My schedule for today” or “Today’s plan” should be only known to the authorized person and not any random persons who asks the assistants. So, we planned on building a voice assistant to remove the above flaw of the currently available assistants, by adding a biometric security, i.e. the face recognition system.

1.4. Scope

With growing the use of data and almost being completely dependent on data, security is needed more than before and what could be better biometric security feature. A facial recognition system is a computer application for automatically identifying or verifying a person. One of the way is to do this is by comparing selected facial features from the image and a facial dataset. It is typically used in security systems and can be compared to other biometrics such as fingerprint or eye iris recognition systems. In this project we focus on 3-D facial recognition system and biometric facial recognition system.

Chapter 2

LITERATURE SURVEY

2.1. Summary of Prior Works

2.1.1. Voice Assistants

a) Aditya Sinha et al. conferred a virtual voice based mostly personal intelligent assistant for visually disabled persons. The project could acknowledge and reply to what user could say in a good and efficient manner via voice, similar to having a spoken language. The main focus was on the development of informal agent and speech recognition module which would be able to work offline and perceive the Indian accent. The approach included the employment of Java library Sphinx-4, MaryTTS and neural networks to enter the training capabilities.

b) Moreover, in 2017 Othman planned paper on Voice Controlled Personal Assistant Victimization Raspberry Pi, the project showed the implementation of a Voice Command System as Associate in Nursing Intelligent Personal Assistant (IPA) which would perform various tasks or services for a private victimization Raspberry Pi as a main hardware to implement this model that works on the first input of a user's voice. Bibek Behera had planned a model for a private assistant to ease out the work done by humans for tasks like booking tickets, ordering food, etc. Chappie uses linguistic communication process (NLP) to analyse chats and extracts intent of the user. Then it uses this info and AIML (Artificial Intelligence Mark-up Language) to form a conversation with the user.

2.1.2. Gender Recognition

A rich amount of the literature is already existing on the topic of the gender classification. It's not possible to organize all the previous methods into the single ubiquitous taxonomy in this current paper; however, we have provided a quick overview of this a how researcher previously approached to gender classification.

The earlier methods which are addressed the gender classification problems are known as the appearance-based methods. In this appearance based methods, features are extracted from face image—considering that the face as a one-dimensional feature vector & then the classification tool is used. With some earlier researcher extracted pixel intensity value as well & then fed these value to the classifier. The pre-processing step included face alignments,

image re-sizing, & illumination normalization. The sub-space transformation is also performed either to reduce dimensions or to explore parts of the underlying structure of image raw data. Basically, the classification is mainly performed through the binary classification strategy. The mostly used classifier for automatic gender classification is the support vector machine; other classifiers applied are included decision trees, neural networks, and AdaBoost. For more than detailed information about gender classification methods.

Recently, new gender classification algorithm is proposed in some papers. Authors of that paper performed tests on the two databases FEI and a self-built database. Different kind of texture features are extracted from the face images. The texture features are extracted from the three discriminating levels which includes global, directional, and regional. The kernel-based support vector machine was used later on for the classification stage.

The other methods which were used for the gender classification are known as geometric methods. These model extract the facial landmark information from the face images & build a model based on the landmarks information. Geometric models maintain certain geometric relationship between different face part. These model discard the facial texture information in the whole modelling process.

Deep Convolutional Neural Networks (CNNs) showed an outstanding performance for the various image recognition problem. The CNN based methods were applied to both features extraction as well as the classification algorithm for automatic gender classification. Hybrid system for the gender & age classification was presented. Features were extracted through the CNNs, & an extreme learning machine (ELM) used for the classification. This hybrid model is known for the ELM-CNNs in the literature. The ELM-CNNs evaluated on two public databases, MORPH-II & Adience. The ELM-CNN is best algorithm for performing on the gender classification thus far.

2.1.2.1. Gender Classification

A detailed survey of the gender classification method that can be found. Here we quickly survey relevant methods. Early methods for the gender classification are used neural network trained on the small set of the near-frontal face images. In the combined 3D structures of the head & image intensities are used for classifying gender. SVM classifiers are used, applied directly to image intensities. Rather than using the SVM, used the AdaBoost for the same purpose, here again, applied to image intensities. Finally, the viewpoint-invariant age & gender classification was presented. Different works were done on the facial gender detection

which introduce the unique result with their performance rate for the different database. Those methods basically rely on the following causes: what were the basis for face features extractions? How will be done analysis of extracted feature & result? What type of the sample databases have been taken? Then after gender detection process is being carried out. H. D. Vankayalapati has contributed in his work based on Support Vector Machines(SVM) algorithm for the feature classification using MATLAB. Facial edge has been carried out using the Laplace of Gaussian filter to determine the landmark position. GTAV face database is being used for the verification of the input data. The limitation of the work is classification may differ with the human race. Hence to eliminate the limitation of race & ethnicity Elham Arianasab presented work using Neural Network-based classification algorithm for the gender diagnosis & reliability is mainly based on the pixel value and geometric facial features.

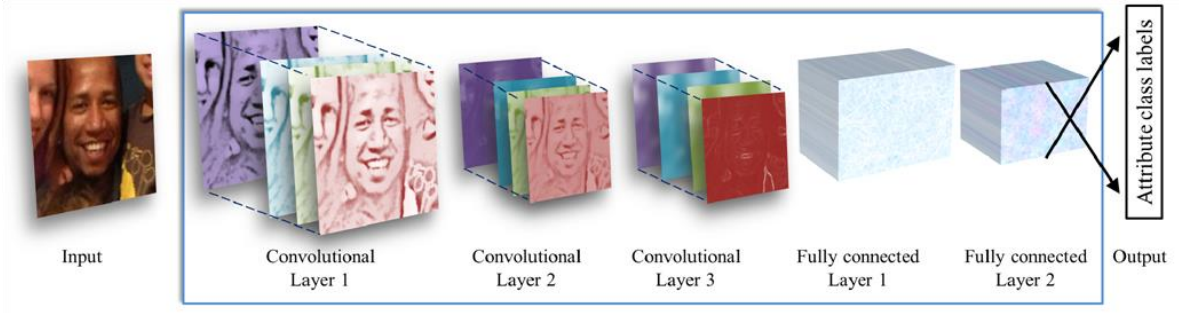


Figure 2.1 – Gender Classification using CNNs

2.1.3. Face Recognition and Detection

Face detection is technology that verify the locations & sizes of external body part discretional image. The facial features are detected and also the alternative objects like trees, buildings and bodies etc. are neglected from the digital image. It is thought to be a specific case of object-class detection, wherever the task is finding the placement and sizes of all objects in a picture that belong to a given category. Face detection, is thought to be a lot of general case of face localization. Face localization, the task is to search out location and size of the glorious variety of faces. Primarily there have been 2 variety of approaches to sight facial half within the given image i.e. feature base and image base approach. Feature base approach tries to extract choices of the image and match it against the information of the face options. Whereas image base approach tries to induce best match between coaching and testing pictures.

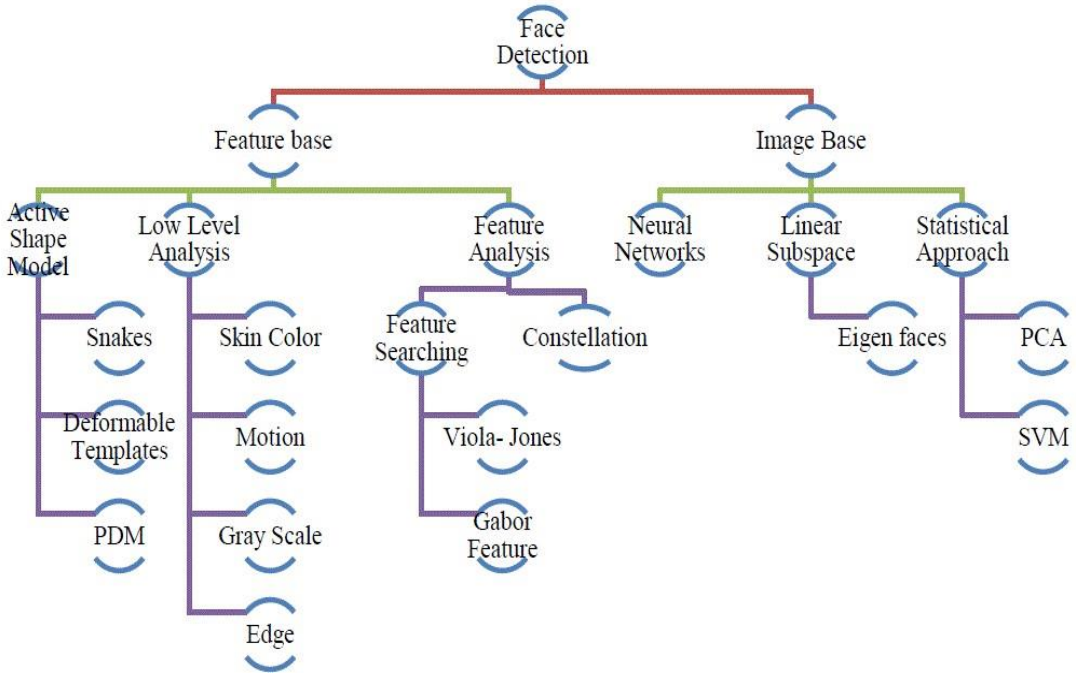


Fig 2.2-Detection methods

2.1.3.1. Face Tracking

Object pursuit is outlined as keeping a trace on a selected quite object. during this paper as we have a tendency to square measure in the main concentrating on face, we have a tendency to track human faces supported the given input options. Continuous pursuit makes us leave the issues like illumination, variation in create etc. aside. Here pursuit of human faces during a video sequence is finished and additionally live video pursuit employing a digital camera is finished.

2.1.3.2. Kanade Lucas Tomasi (KLT) algorithm

Kanade Lucas Tomasi algorithmic program is employed for feature chase. it's a preferred one. KLT algorithmic program was introduced by filmmaker and Kanade and their work was later extended by Tomasi and Kanade. This algorithmic program is employed for police work scattered feature points that have enough texture for chase the desired points during a sensible customary. Kanade-Lucas-Tomasi (KLT) algorithmic program is employed here for chase human faces unceasingly during a video frame. This technique is accomplished by them finding the parameters that permit the reduction in similarity measurements between feature points that area unit associated with original translational model. first during this algorithmic program, we have a tendency to calculate the displacement of the half-track points from one frame to a different frame. From this displacement calculation it's straightforward to reason

the movement of the top. The feature points of a person's face area unit half-track by mistreatment optical flow hunter. KLT chase algorithmic program tracks the face in 2 easy steps, first it finds the traceable feature points within the 1st frame and so tracks the detected options within the succeeding frames by mistreatment the calculated displacement.

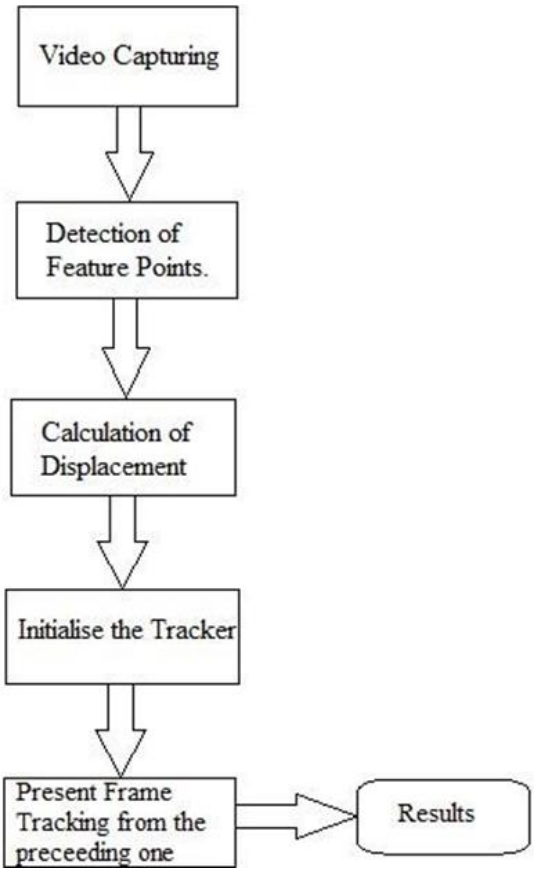
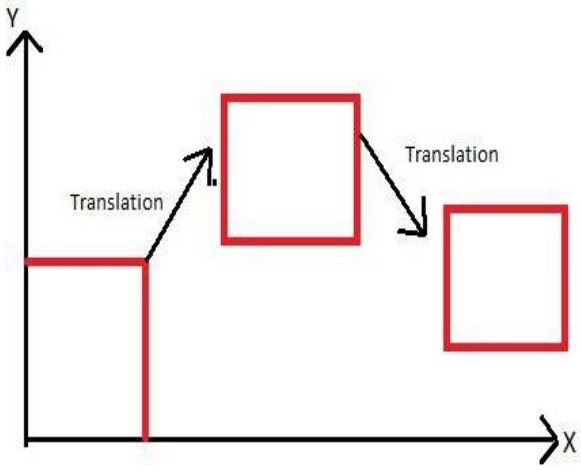


Fig 2.3- KLT Algorithm

2.1.3.2.1. Translation

In this rule foremost it detects harris corners within the first frame. So then optical flow it continues to find the points by computing the motion of the pixels of a picture. For every translation motion of the image optical flow is computed. The harris corners are detected by linking the motion vectors in consecutive frames to get a track for every harris point. Simply to not lose the track of the video sequence we tend to apply harris detector at each ten to fifteen frames. This can be nothing however ensuring by checking the frames periodically. During this method new and previous Harris points area unit half-track. Here during this paper, we tend only 2-D motion i.e. translation movement.

Let us assume that the initially one amongst the corner point is (x,y). Then the next frame, if it's displaced by some variable vector (b1, b2,.bn), the displaced corner purpose of the frame is the total of the initial purpose and displaced vector. The coordinates of the new point are x=x+b1 and y=y+b2. So, the displacement currently should to be calculated with reference to every coordinate. For this we use warp function that may be a function with coordinates and a parameter. it's denoted as W(x;p) = (x+b1;x+b2). The warp function is to estimate the formation.



2.1.3.3. Face Detection

Detecting face is a computer technology which will let us know that the location & size of human faces. This will help in getting the facial feature & avoiding other object & things. In this present situation human face perception is a biggest research area. Basically about detecting a human face through some trained features. Face detection is preliminary step for many other applications such as face recognition, video surveillance etc.

Face detection involves separating image windows into 2 classes; one containing faces (turning the background (clutter). It's tough as a result of though commonalities exist between faces, they will vary significantly in terms mature, skin colour and facial features. the matter is additional sophisticated by differing lighting conditions, image qualities and geometries, further because the chance of partial occlusion and disguise. a perfect face observes or would so be ready to detect the presence of any face below any set of lighting conditions, upon any background. The face detection task may be countermined into 2 steps. the primary step may be a classification task that takes some whimsical image as input and outputs a binary price of affirmative or no, indicating whether or not there are a unit any faces gift within the image. The second step is that the face localization task that aims to require a picture as input and

output the placement of any face or faces inside that image as some bounding box with (x, y, width, height).

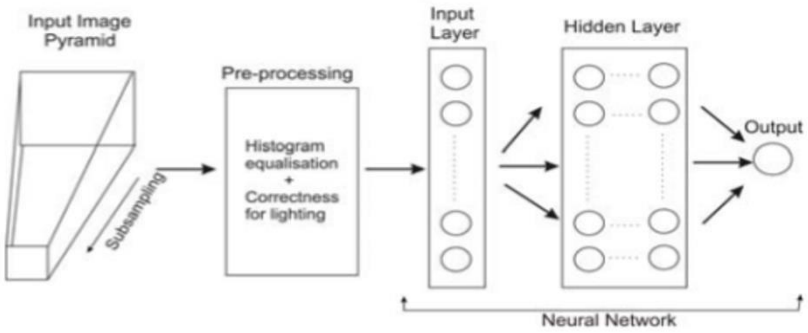


Figure 2.4 –A generalised view of Face Detection Algorithm

2.1.3.3.1. Viola-Jones Algorithm

This algorithm helps us to discover options of a face in any particular frame of a video sequence. This can be the primary object detection framework which gives a competition to real time detection rates. Paul Viola and Michael Jones are the one those who introduced this algorithm. They created this algorithm mainly by the issue of the face detection. There are four steps that have to be followed to discover a face. Firstly, we tend to train the system with the haar features. Haar features are a form of rectangular boxes that are black and white.

Haar features are easy rectangular feature that is difference of the sum of pixels of areas inside the rectangle. This rectangle may be at any position of the frame and may scale the image. This changed feature set is named 2-rectangle feature. Every feature kind will indicate the existence or the absence of sure characteristics within the frame, like edges or changes in texture.

These haar features are applied to work out the facial features. The Black half is used to discover nose feature of a person's face because the black coloured part defines the presence of a nose that is located at the middle of the face. And also the Figure-3 (e) is named a four Rectangle feature. Wherever the black part is denoted as +1 and also the white part is denoted as -1. The result's calculated by subtracting the sum of pixels underneath the white rectangle from the sum of pixels under black rectangle. At first some threshold is taken for specific features. Average sum of every black and white is calculated. Then distinction is checked with threshold. If the worth is on top of or matches with the edge, then it's detected as relevant feature.

2.1.3.3.2. Integral Image Test

The integral image part is to add all the pixels of a selected box to its left and above ones. The four corner values of area that are to be calculated. This makes avoid summing of every element within the region. This integral image conversion method is introduced simply to hurry up the method in calculating pixels.

The calculation of the sum of pixels of the part D within the below figure is (1+4) (2+3) i.e. [A+(A+B+C+D)] - [(A+B+A+C)] which supplies D. The authors defined the bottom resolution of the detector to be 24x24. In different words, each image frame to be divided into 24x24 sub- windows, and feature are extracted in any respect possible locations and scales for every such sub-window. This leads to an exhaustive set of rectangle features that counts over a 160,000 feature for one sub-window.

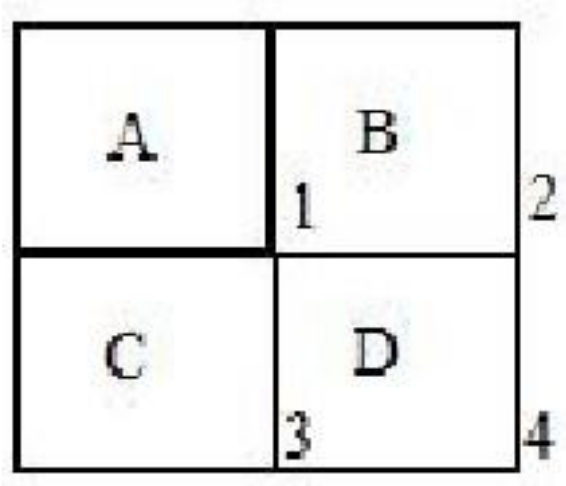


Figure 2.5- Integral Image Classification

2.1.3.3.3. ADA-Boost

It is a process which is used to find out the relevant & irrelevant features. It uses weak classifier & weight to form the strong classifiers. It finds the single rectangular features & thresholds which are the best to separate the non-faces & faces in training example in term of weighted error. It first starts with the uniform weight while training. Next it evaluates the weighted errors for every features and picks the best one. Than we re-evaluate the example where the incorrect classifiers will have more weight & correct classifiers will have less weight. Finally, classifier will contain the combination of the correct classifiers which are

having mainly less weight. Which reduce the computational time of non-faces to be discarded.

2.1.3.3.4. Learning Classification Functions

The complete set of the features is quite large which are 160,000 features per single 24x24 sub-window. By computing a single feature can be completed with only a few simple operation & evaluating the entire set of features which is still extremely expensive and cannot be performed by the real-time applications. Viola & Jones assumed that the very small number of extracted feature can be mainly used to form an effective classifier for the detection of face. Hence, the main challenge is to find those distinctive feature. They decided to use the AdaBoost learning algorithm for the feature selection mechanism. In the original form, AdaBoost is mainly used to improve classification results of the learning algorithm by combining a collections of weak classifier to form the strong classifiers. This algorithm starts with equal weights for all the examples. In every round, the weights are updated so that misclassified example receives more weight. By drawing this analogy between the weak classifiers and features, Viola & Jones decided to use the AdaBoost algorithm for the aggressive selection of some small number of good feature, which nevertheless have significant varieties. Theoretically, the weak learning algorithm is restricted to the set of classification function, which of every is dependent on the single feature. And the weak classifier h (x, f, p) is then defined for a sample x (i.e. 24x24 sub-window) by the feature f, a threshold, and the polarity p which is indicating the direction of the inequality. The main advantage of the AdaBoost over its competitors is speed of learning. For every feature, the example is sorted based on the feature value. The optimal threshold for this feature can be then computed in the single pass over these sorted list.

2.1.3.3.5. Cascading

The step is mainly introduced to speed up the process & give an accurate result. This step constitutes of several stage where every stage consists of a strong classifier. Every features are grouped into some stages. It detects the faces in the frame by sliding a window over a frames. Whenever an input is given it checks for the certain classifier in the first stage and then so on. But it is passed to successive stage if & only if it satisfies the preceding stage classifier.

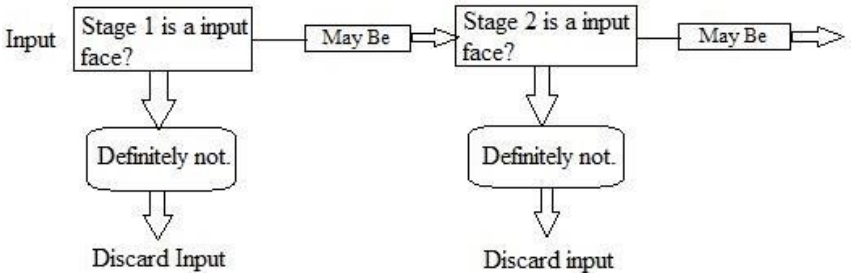


Figure 2.6- Cascading

2.1.3.4. Constellation Method

All strategies mentioned to this point are ready to track faces however still some issue like locating faces of assorted poses in advanced background is actually troublesome. To scale back this problem investigator, form a group of facial features in face-like constellations using additional sturdy modelling approaches like applied math analysis. Numerous varieties of face constellations are planned by Burl et al. They establish use of statistical shape theory on the features detected from a multiscale Gaussian derivative filter. Huang et al. additionally apply a mathematician filter for pre-processing in a very framework supported image feature analysis. Image Base Approach.

2.1.3.4.1. Neural Network

Neural networks are gaining so much more attention in many patterns recognition problem, such as OCR, object recognition, & autonomous robot driving. Since the face detection can be treated as the two class pattern recognition problems, and various neural networks algorithm have been proposed. Advantage of using neural network for the face detection is the feasibility of the training a system to capture complex class conditional density of the face pattern. However, there is demerit in the network architecture has to be extensively tuned to get exceptional performance. In days many hierarchical neural networks are proposed by Agui et al. The first stage having two parallel sub network in which inputs filtered intensity values from an original image. Inputs to the second stage network consists of the output from the sub network & extracted feature value. The output at second stage shows the presence of the face in the input regions. Propp and Samal developed one of the earliest neural networks for face detection. Their network consists of four layers with 1,024 input units, 256 units in the first hidden layer, eight units in the second hidden layer, and two output units. Feraud &

Bernier presented the detection method using auto associative neural network. The idea is basically based on the auto associative network with five layer which is able to perform a nonlinear principal component analysis. One auto associative network is used to detect the frontal view faces and another one is used to detect the face turned up to 60 degrees to the left & right of the frontal view. After that Lin et al. presented the face detection system using probabilistic decision-based neural network (PDBNN). The architecture of PDBNN is similar to the radial basis function (RBF) network with modified learning rules & probabilistic interpretation.

2.1.3.5. Statistical Approach

2.1.3.5.1. Support Vector Machine (SVM)

SVMs were first introduced Osuna et al. for face detection. SVMs work as the new paradigm to train polynomial functions, neural network, radial basis function (RBF) classifiers. SVMs works on the induction principles, called structural risk minimizations, which targets to minimize the upper bound on expected generalization errors. The SVM classifier is the linear classifier where the separating hyper plane is chosen to minimize an expected classification error of the unseen test pattern. In Osuna et al. developed the efficient method to train the SVM for large scale problem, and are applied it to face detection. Based on two test sets of 10,000,000 test patterns of 19 X 19 pixels, their systems have slightly lower error rates & runs approximately 30 times faster than system by Sung and Poggio. SVMs had also been used to detect faces and pedestrians in the wavelet domain.

2.1.3.5.2 Digital Image Processing

Digital Image Processing Methods stock from two principal application areas:

1. Improvement of the pictorial information for the human interpretation.
2. Processing of the scene data for the autonomous machine perception.

In the second application area, it focuses on procedure for the extracting images information in the form which are suitable for the computer processing.

Examples include automatic character recognitions, industrial machine visions for the product assemblies and inspections, military recognizances, automatic processing of fingerprints etc.

Image:

An image refer an 2D light intensity functions f(x, y), where(x, y) denote the spatial coordinate and the value of the f at any point (x, y) is proportional to brightness or gray level of the images at that point. The digital image is an image f(x, y) that has discretized both in the spatial coordinate & brightness. The element of such a digital array are called as image elements or pixel.

A simple image model:

To be suitable for the computer processing, the image f(x, y) must be digitalized both spatially & in amplitude. Digitization of the spatial coordinate (x, y) is called as image sampling. Amplitude digitization is called as gray-level quantization.

The storage & processing requirement increased rapidly with the spatial resolution & the number of gray levels.

Example: 256 gray-level image of the size 256x256 occupies the 64k byte of memory.

Type of Image processing

- Low Level Processing
- Medium Level Processing
- High Level Processing

2.1.3.5. Face Recognition

There are two prevailing which proceeds towards to the face recognition problem: Geometric which is feature based and photometric which is view based. As researchers interest in face recognition continued, many different algorithms were developed, three of which have been well studied in face recognition literature.

These algorithms can be divided into two main approaches:

1. **Geometric (feature based):** It is based on the geometrical relationship between the facial marks, or in other words we can call it as the spatial configuration of facial features. Which means that the geometrical features of the face are such as the nose, eyes and mouth are first located and then the faces are classified on the basis of the various geometrical distance & angle between the features.

2. **Photometric stereo (view based):** It used to recover the shape of an object from a number of images taken under different lighting conditions. The shape of object recovered is described by gradient map, which is done up of an array of surface normal.

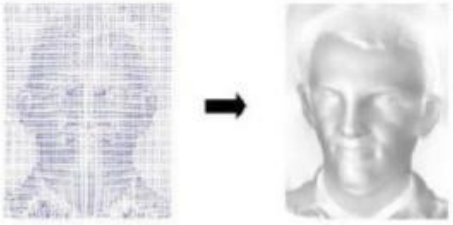


Figure 2 -Photometric stereo image.

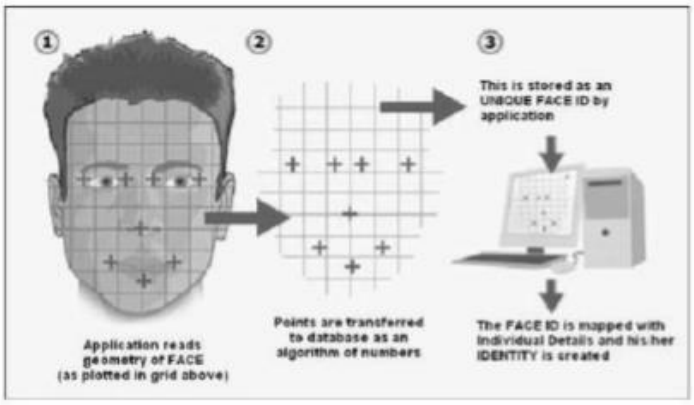


Figure 3 - Geometric facial recognition.

Popular recognition algorithms include:

A. Principal Component Analysis (PCA)
B. Linear Discriminate Analysis,
C. Elastic Bunch Graph Matching Using Fisher Face Algorithm,

2.2. EXISTING SYSTEM

Google Assistant in Android, Bixby in Samsung devices, Sir in Apple devices respectively, are great examples of general purpose Voice Assistant which are present on smartphones, smart speakers and other smart devices. All of these Voice assistants help us with general-purpose things like setting the alarm, scheduling events, making calls, launching apps, playing music, amongst other things. Google Assistant also has a feature called App Actions which can trigger actions inside specific apps using deep links. Recently, Alexa announced the same with Alexa on apps. Unfortunately, Siri doesn't have any such feature currently and

has poor performance compared to both Alexa and Google Assistant. Increasingly, Google Assistant and Alexa are shifting strategy and focusing on doing tasks inside apps encroaching on the niche carved by in-app Voice Assistants.

2.3. PROPOSED SYSTEM

With all the current voice assistants available such as Alexa, Google assistant and various others, there is a little security flaw. As the current voice assistants, all being voice controlled, none of them identifies the user who commands it. This issue of not identifying the user may rise to a problem of giving away the data to the user who shouldn't have it. The proposed system is to implement a voice assistant with the face detection, face recognition, gender recognition and various models using various supervised and unsupervised datasets and test its usability in a voice assistant. So, in general words, we planned on building a voice assistant to remove the above flaw of the currently available assistants, by adding a biometric security, i.e. the face recognition system.

Chapter 3

SOFTWARE AND HARDWARE REQUIREMENTS

3.1. Functional Requirements

1. Faces of the users to train the face recognition model.
2. The gender dataset to train the gender recognition model.
3. The system should be able to produce required output based on the input provided.
4. There should be interaction between user and the system.
5. Complete analysis and prediction of the output based on the model.
6. Ease of usability and maintainability to user.

3.2. Non-functional Requirements

We aim at attaining the following attributes of quality in our project:

1. Ease of use

   As our target population are the citizens of our society, the project developed must be easy to use requiring minimum assistance in interacting with the environment keeping in mind the physical limitations.

2. Data Integrity

   We need to ensure that data is accurate, complete and repeatable which is an utmost requirement for running the suggestive Deep learning algorithm to give accurate results.

3. Maintainability

   The error debugging process is also an essential factor as the rendering speed of the environment must be fast so as to produce the required output.

4. Reliability

   The system should propose a reliable outcome as it is concern with the life of people.

3.3. Hardware Requirements

CPU	Intel i3 and above
Memory	5GB and above
RAM	8 GB and above

OS	Windows 10
Camera	Any good resolution(5MP recommended)
Microphone	Should be Available

Table 3.1 Hardware Requirements

3.4. Software Requirements

1. Language- Python 3.7.5
2. IDE- Any Python editor (Pycharm used here)
3. scikit-learn(Version: 0.24.1): used for implementation of algorithms (random forest, SVM and neural networks)
4. Utils
5. Sys
6. Pandas (Version: 0.24.2)
7. Numpy (Version: 1.18.1)
8. Ipaddress - IPv4/IPv6 manipulation library
9. Requests (Version: 2.23.0)
10. Re - Regular expression operations
11. SpeechRecogniton (Version 3.8.1)- for speech recognition
12. opencv-python (Version: 3.4.2.16)
13. playsound
14. pyjokes
15. and various other modules were imported as per requirements

3.5. Cost Estimation

Our project does not require us to buy any hardware because all the hardware needs can be satisfied with our laptop computers. All the software required for our project are open-source, hence they are all free. So apart from the cost for publishing research paper and printing reports, it is costless.

Chapter 4

SYSTEM DESIGN

4.1. Architecture

The proposed system is to implement a voice assistant with the face detection, face recognition, gender recognition and various models using various supervised and unsupervised datasets and test its usability in a voice assistant.

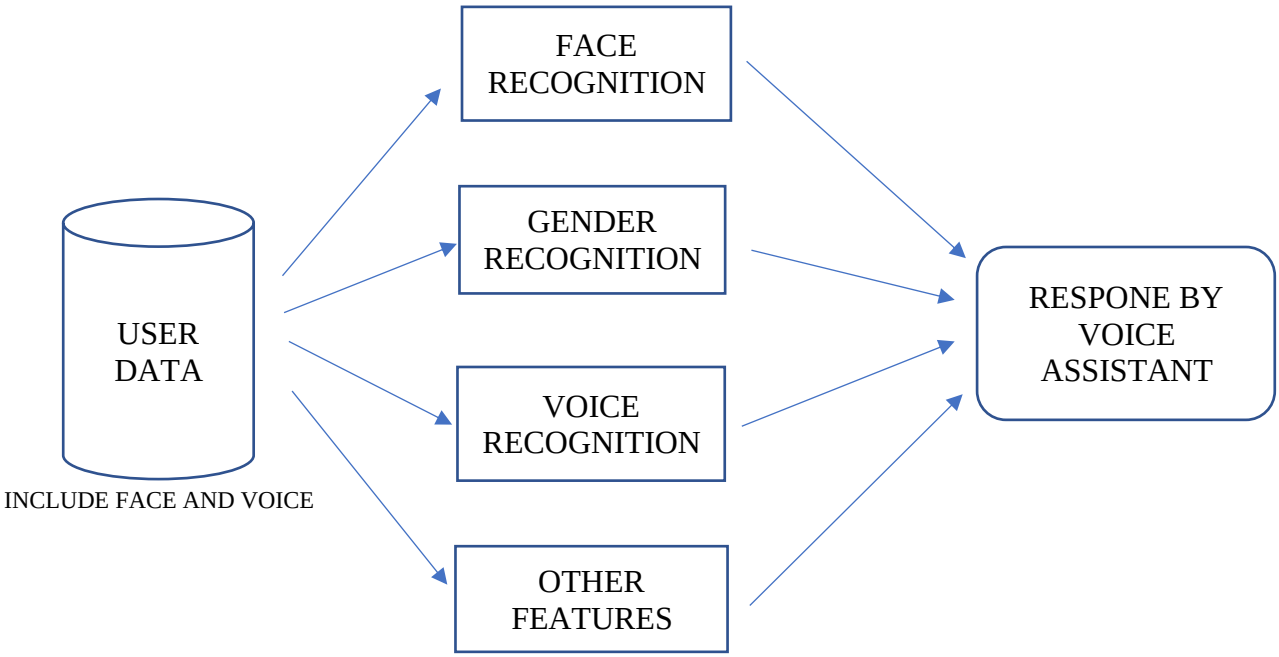


Figure 4.1- High Level Design

The above figure depicts the high level approach adopted to the problem statement through the training of various Models of face recognition and gender recognition with our training dataset and aggregating the results to implement a good voice assistant.

4.1.1. System Architecture

In the Figure 4.2, high level graphical representation of architecture of system is portrayed. Dataset was optimized through extraction of features. After training and testing the proposed models with optimized dataset and well implemented and integrated for deployment.

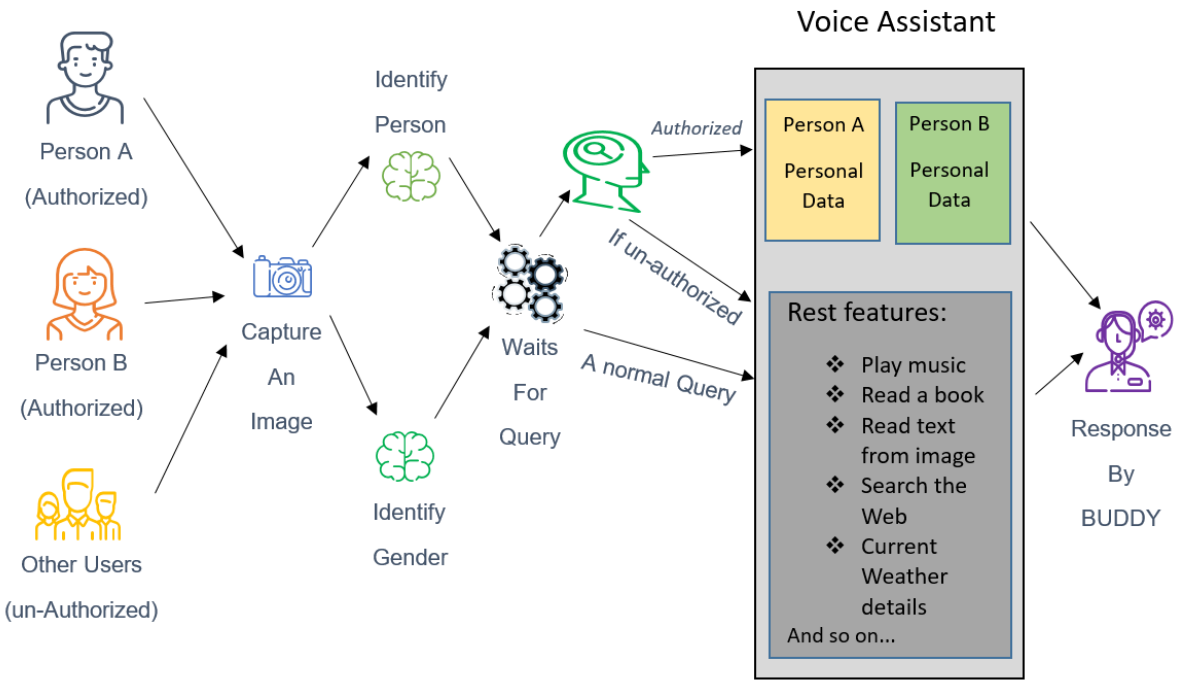


Figure 4.2- System Architecture

4.1.2. Use-case Diagram

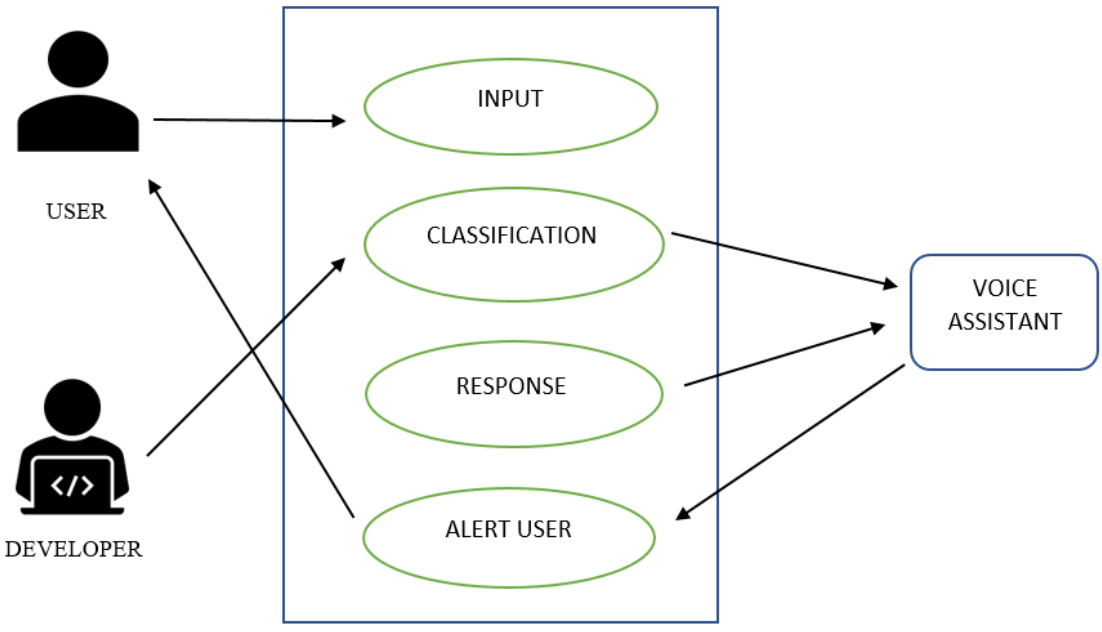


Figure 4.3- Use-case Diagram

The use case diagram for a normal execution is represented in the above figure. A user (who has installed the application) opens the application, in which the face of the user is an input to the project. The developer handles the accuracy rates to classify whether the person is authorized or not. The face recognition system embedded in voice assistant classifies the user (authorized or unauthorized) and sends the result back to the interface. The interface has the confidence and recognized user name along with their gender, and if any personal question is asked then it will be sent a text message as an alert.

4.2. Detailed Design

4.2.1. Face Recognition

4.2.1.1. System Architecture

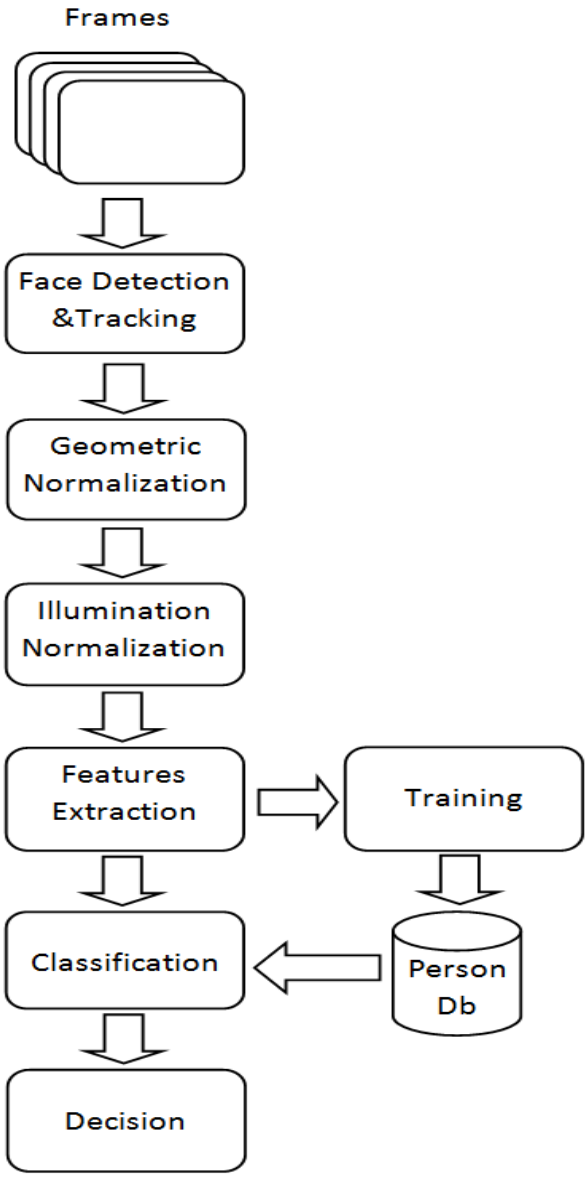


Figure 4.4- System Architecture of Face Recognition System

4.2.1.1.1. Geometric normalization

It is very important to detect and track the faces in all conditions and variations. When comparing local regions between faces, an image registration step must be performed so corresponding facial features are synchronised.

Simple geometric normalization usually involves bringing the faces to a standard size and rotating them in-plane in order to bring the eyes on the same horizontal line. Figure 2 shows some face samples before and after applying the geometric normalization.

More complex normalization scenarios (Corcoran et al., 2006b) can use 3D face models to rotate the face in the out-of-plane space to have identical orientation (i.e only frontal faces). This will have a higher computational requirement and could only be used when there is enough processing power. Figure 3 shows an example of the output of this complex normalization which can help recognition for large pose variations.

All other processing steps applied after geometric normalization should have the same effect on each face.

4.2.1.1.2. Illumination normalization

If we can control the image capturing environment and impose strict requirements regarding lighting conditions (i.e. control access), recognition accuracy can be improved. In most scenarios where video face recognition is employed, the variations in lighting conditions when the faces are captured can range between dark and bright extremes. The profile face samples for each person to be recognized are captured in very different conditions with still images than those used in video imagery. A pre-processing algorithm should be used to minimize the effect of the lighting conditions when capturing the video images.

4.2.1.1.3. Feature extraction

Together with the useful information that can be used to differentiate between individuals, the face images described by the pixel values contain redundant information and information that can be ignored in the classification stage. By extracting only, the useful information in this step we improve the accuracy of the recognition and also lower the storage requirement for each face.

4.2.1.1.4. Classification

In the case of still image recognition, the system makes a decision if the test face belongs to one of the people in the database and if so, which one (based on comparing the features computed in the previous step for test faces and a database of people).

Simple classification algorithms like distance between feature vectors are preferred because of their simplicity and speed. More complex learning algorithms can be used if there are enough computation resources.

The classification algorithm is divided into two stages:

1.Training- Prototypes are constructed for each person in the database. The prototypes can be built from single or multiple face samples. Using multiple samples improves the quality of the prototype. The prototype can be represented by a series of feature vectors (such as distance-based classification) or can be represented by statistical models trained with multiple samples (such as learning-based classification algorithms).

2.Testing- Test samples are compared with each person prototype and similarity scores are computed. A decision is made using these similarity scores and the history of previous scores.

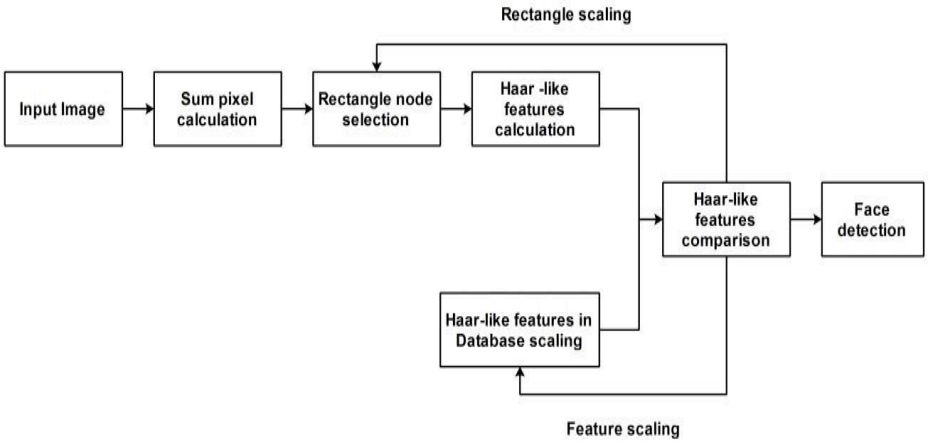


Figure 4.5- Diagram of the Haar-face detection algorithm

The operations of face detection are explained as follows:

(a) Input Image: This is for taking an image of the person that we want to detect his/her face.

(b) Sum of the pixels: We should calculate the sum of the intensity of values of an image by the following formulas:

$$ii(x, y) = \square_{x'<=x,\ y'<=y}\ i(x', y') \quad (1)$$

Where: i(x, y) is intensity of the grey scale image at the pixel (x, y).

(c) Rectangle node selection: After indication of the four points of the rectangle (A, B, C and D). We make the summation of the intensity pixels of the rectangular area ABCD which can be calculated from the Equation. The illustration parts of Eq.(2) is indicated in Figure.

$$\text{Sigma}(x,y)\ \text{Sigma}\ ABCD\ \ I(x, y) = ii(D) + ii(A) - ii(B) - ii(C) \quad (2)$$

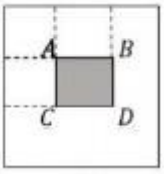


Figure.5 Rectangle area of ABCD

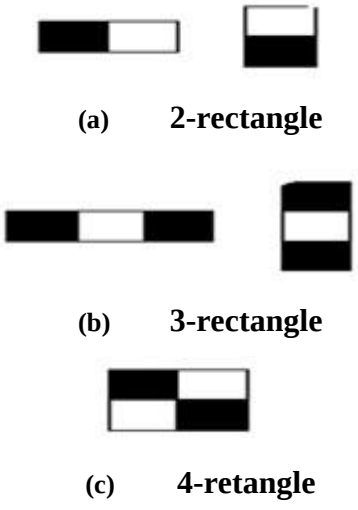
(d) The Haar like features calculation: Examples of haar-like features are shown in Figure including 2-rectangle filters, 3-rectangle & 4-rectangle. The features which can be applied to the original image from the Eq.(3)

f=[ii(D)+ii(A)-ii(B)-ii(C)]—[ii(F)+ii(C)+ii(D)-ii(E)] (3)

(e) The Haar like Features comparison: To test the features & train the classifiers, we use Ad boost as a classifier & it can be defined as:

$$h(x, f, p, \text{theta}) = \begin{cases} 1 & if\ pf(x) < p^{\text{theta}} \\ 0 & otherwise \end{cases} \quad (4)$$

where: f is selected feature from the set feature in Figure.5, p is polarity, theta is the chosen threshold & x is training sub window of size 24x24 pixel. Figure.
illustrates face detection step in real examples.



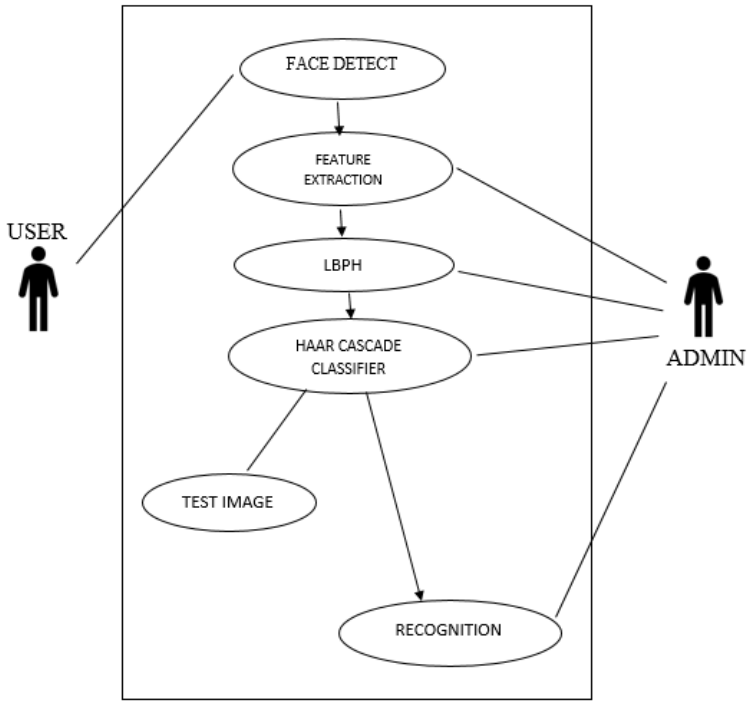
(a) 2-rectangle

(b) 3-rectangle

(c) 4-retangle

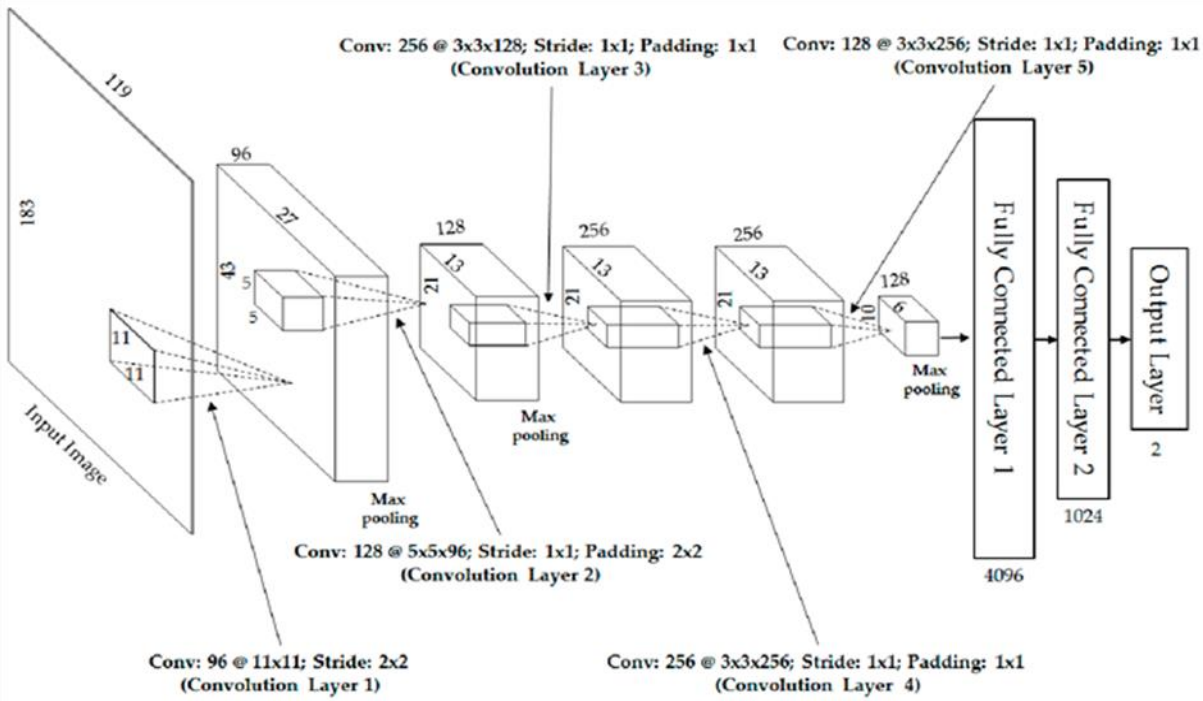
Figure 4.6- the Haar-like features face detection

4.2.1.2. Use-case Diagram

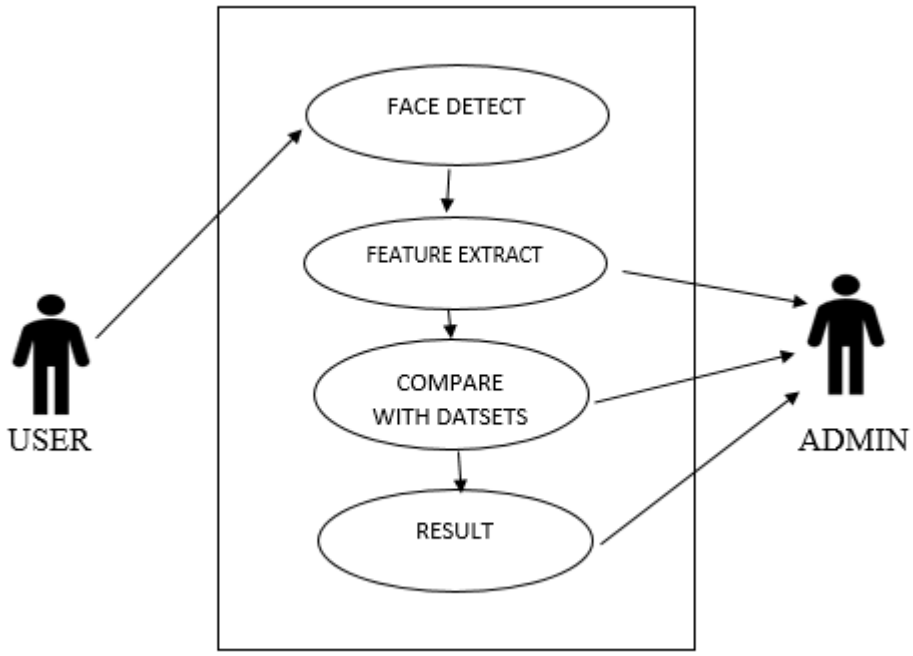


4.2.2. Gender Recognition

4.2.2.1. System Architecture



4.2.2.2. Use-case Diagram



4.2.3. Voice Recognition: -

4.2.3.1. System Architecture

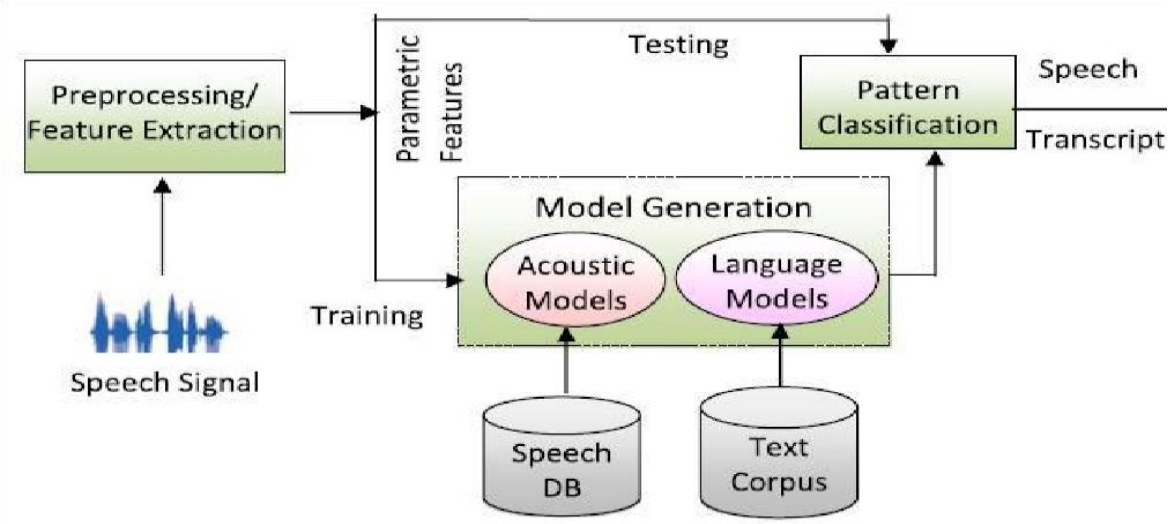


Figure 4.7- System Architecture of Speech Recognition

In the above figure 4.4, a detailed depiction of system architecture has been portrayed. For the speech recognition, a pre-determined module named “pyttsx” was used. It takes the voice as an input from the source (via microphone) and converts into characters and further processing is done.

4.3. Class Diagram

In our project, we have implemented the classifier and GUI for the system with Python. We didn’t use any classes and simply connected the files through function calls. Hence, we have not included class diagram in the report to avoid misinterpretation.

4.4. Sequence Diagram

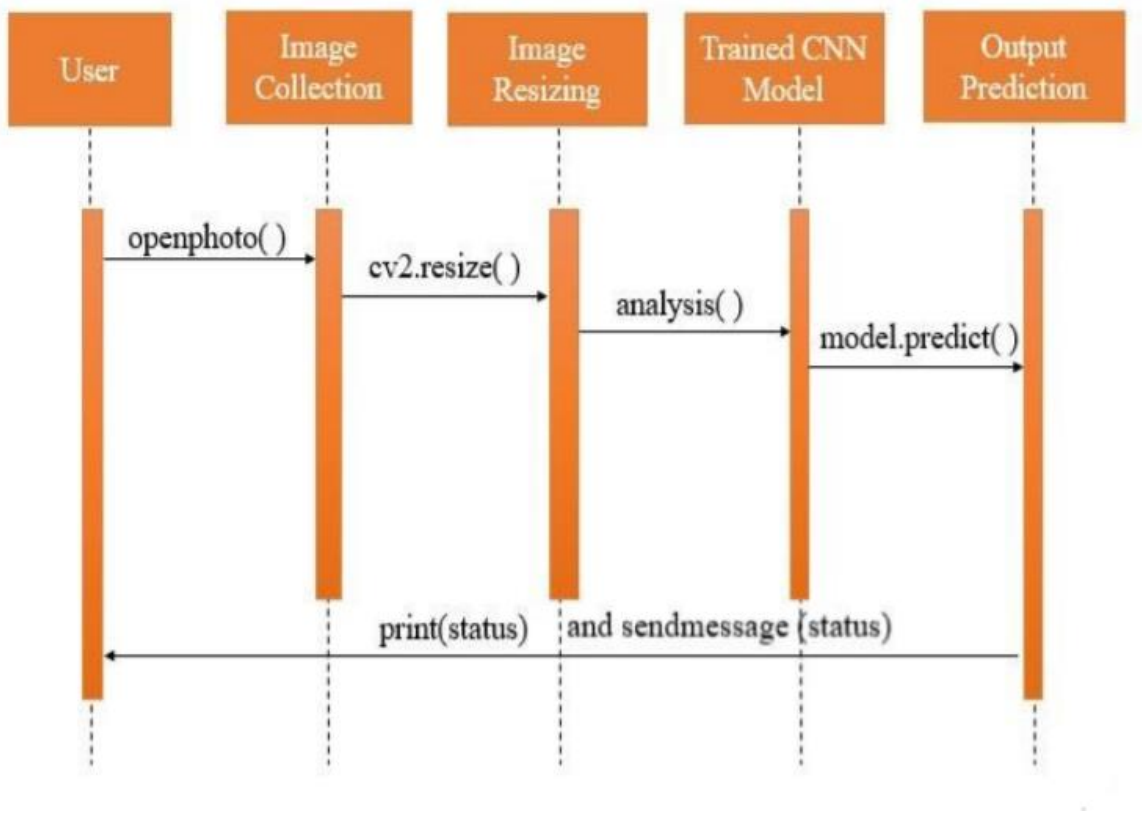


Figure 4.8- Sequence Diagram

UML Sequence Diagrams are interaction diagrams that detail how operations are carried out. They capture the interaction between objects in the context of collaboration. Sequence Diagrams are time focus and they show the order of the interaction visually by using the vertical axis of the diagram to represent time what messages are sent and when. Sequence Diagrams captures: the interaction that takes place in a collaboration that either realizes a use case or an operation (instance diagrams or generic diagrams) and high-level interactions between user of the system and the system, between the system and other systems, or between subsystems (sometimes known as system sequence diagrams).

Purpose of Sequence Diagram: Model high-level interaction between active objects in a system. Model the interaction between object instances within a collaboration that realizes a use case. Model the interaction between objects within a collaboration that realizes an operation. Either model generic interactions (showing all possible paths through the interaction) or specific instances of an interaction (showing just one path through the interaction).

Chapter 5

IMPLEMENTATION

5.1. Voice Assistant

5.1.1. Proposed Plan

The work started with analysing the audio commands given by the user through electro-acoustic transducer. this may be something like obtaining any data, in operation computer's internal files, etc. this is often AN empirical qualitative study, supported reading higher than mentioned literature and testing their examples. Tests square measure created by programming consistent with books and on-line resources, with the explicit goal to search out best practices and an additional advanced understanding of Voice Assistant.

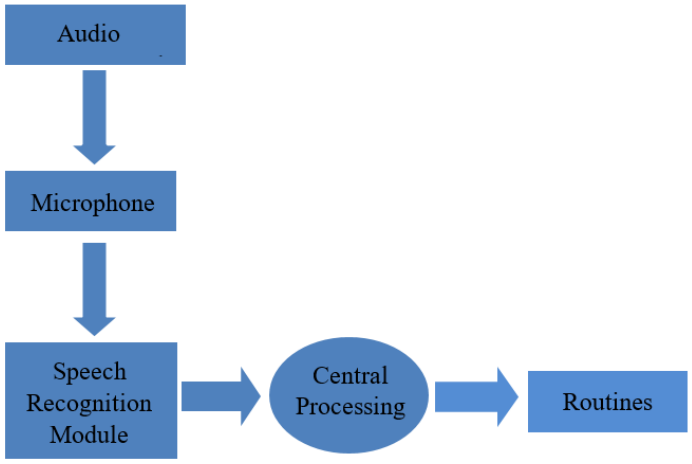


Fig 5.1- shows the advancement of the fundamental method of the voice assistant.

Fig.2 shows the advancement of the fundamental method of the voice assistant. Speech recognition is employed to convert the speech input to text. This text is then fed to the CPU that determines the character of the command and calls the relevant script for execution.

5.1.2. Component and Working

This project aim is to build a personal voice assistant that will make easy for users to use computer with voice command and make task easier. To implement intelligent personal voice assistant, python libraries and Google Speech Recognition API's are used for speech

recognition module and to interpret voice response. Python 3.7 or above, Spyder IDE or Visual studio code or PyCharm, is used for development.

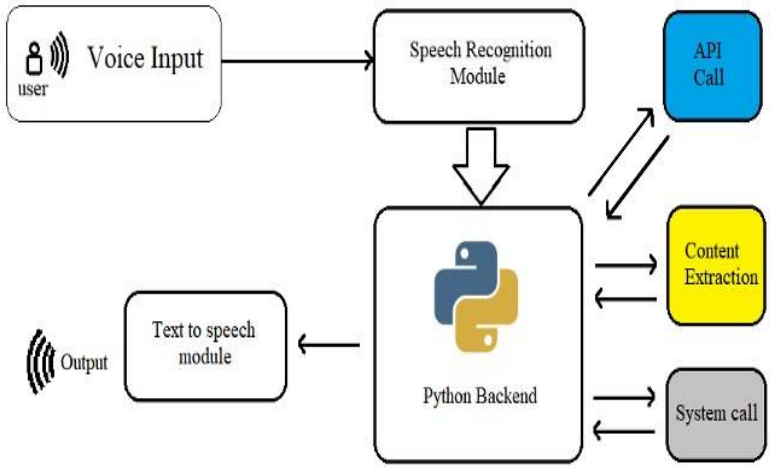


Figure 5.2. Parts of the planned voice assistant

Following at the key parts of planned voice assistants

a) Speech Recognition module. The system uses Google's speech recognition system for changing speech input to text. This module uses the library of speech Recognition three.8.1 for performing arts speech recognition, with support for many engines and genus APIs, online and offline.
b) Python backend get the output from the speech command or the speech output is associate API decision, Context Extraction, and supervisor call instruction then the output is remit to the python backend to offer the specified output to the user.
c) API Calls. It permits 2 applications to speak to every different API is functioning as a traveller that delivers request to the supplier that at requesting it from then delivers the response back.
d) Context Extraction is the task performed of mechanically or robotic extracting structured information from unstructured or semi-structured or each code documents. In majority of the instance this task issues process of human language texts by the strategy of tongue process (NLP).
e) System Calls, is the programmatic methodology during which a laptop program appeal a service from the kernel of the OS it's dead on. It embraces hardware connected assistant and services like access of disk drive, creation and execution of latest processes, and interacting with integral kernel services like method planning. It delivers a necessary interface between a method and also the OS.

f) Text-to-Speech (TTS) engine refers to the power of computers to scan text. The engine converts written language into a sound illustration, then it'll converts the sound illustration to waveforms which are often output as sound.

Intelligent personal voice assistants were meant to control the applications online as well as offline and to analyse the performance of the project the working of Intelligent personal voice assistant with desktop application is shown in figure 3. Opening YouTube with voice command.

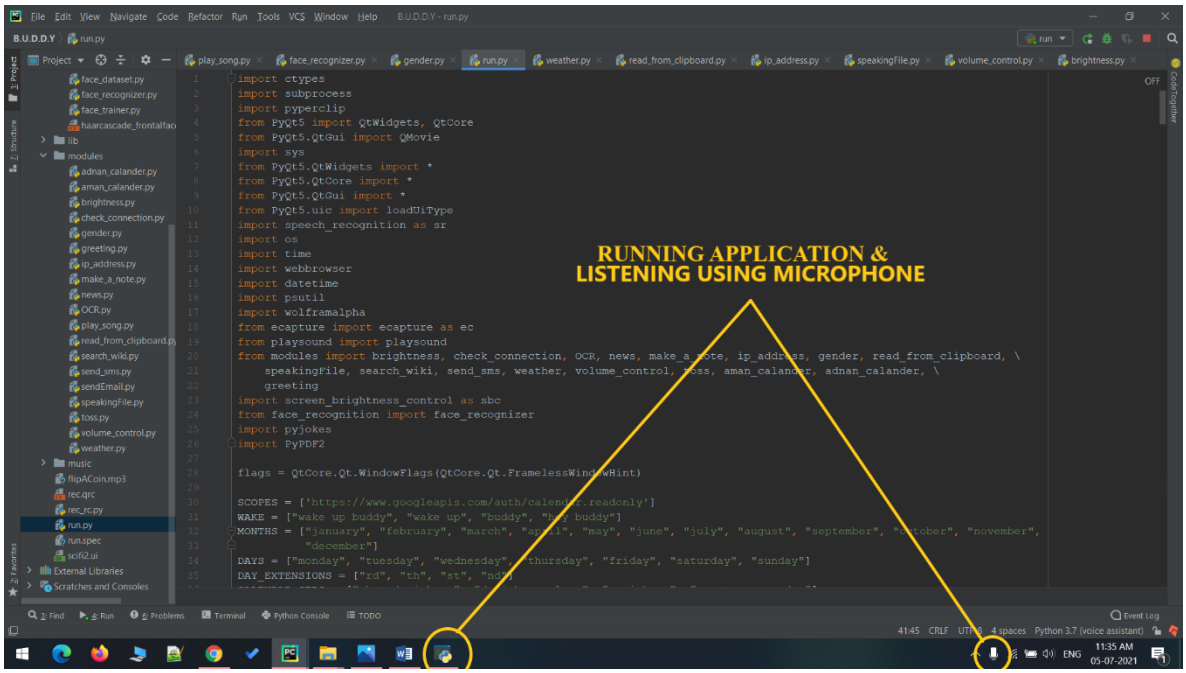


Figure 5.3. Working-Activating Microphone

5.2. Gender Recognition

5.2.1. Face Detection from image

5.2.1.1. Finding faces by color

Since color is liable for providing associate degree economical technique that is computationally effective and strong just in case of depth and partial occasions, it's still being employed. It will work with a mixture of assorted ways like motion and look based mostly face detection. person has skin that is formed of a decent assembly in color area. that's why, once completely different races are thought-about, this technique performs well.

5.2.1.2. Edge-Orientation Matching

This technique may be a native image feature to model objects for finding position and is in a position to detect image in real time. Edge options area unit extracted from a 2nd array of pixels that works because the basic feature for calculation. This is often a gradient primarily based technique which needs convolving the image. The convolution of the image with the filter masks offers two edge strength pictures. Sample of hand-labelled face pictures area unit want to produce face model. The faces area unit cropped, aligned and scaled within the gray level domain. From this set of normalized face pictures a mean face is computed. Vertically reflected versions of every face within the set area unit adjusted to the typical face. Finally, the sting orientation vector field is calculated from the typical face. For face detection, the model is shifted over the image, and at every image position the similarity between the model and also the underlying image patch is calculated. The image is pictured by its orientation field.

Figure 5.4: edge orientation vector field computed

5.2.1.3. Hausdorff Distance

This method is considered as a robust scheme which uses Hausdorff distance for measuring similarity among general face model and probable instances of the object. Hausdorff distance is considered as a metric among sets of two points. Illumination and background is changed robustly by it.

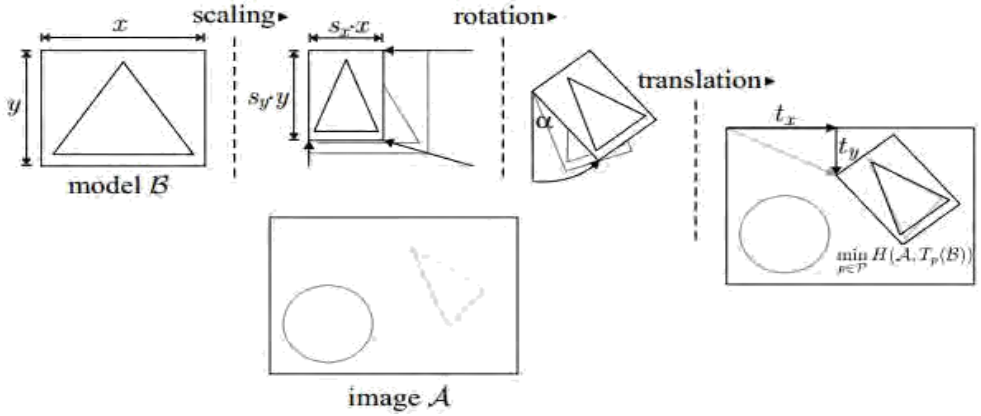


Figure 5.5- Model fitting by translation, rotation and scaling

5.2.2. Pre-Processing of the Image

5.2.2.1. Local Pixel Grouping

Local Pixel Grouping is used for noise reduction because of its elementary nature. Principle Component Analysis (PCA) helps the system to be stable while noise is being cleared. It works generally well. Low variance components are discarded by PCA in the time of conserving Principal Components with Larger variance.

5.2.2.2. Linear Filter

Linear Filter is used for removal of noise using various filters like averaging filters or Gaussian filters, thought it depends on the type of noise. For an example averaging filter is used for grainy type of noise in an image. Here, each pixel gets the average value of neighbour pixels.

Figure 5.6- Linear filtering of "salt and pepper" noise

5.2.2.3. Median Filter

Median Filter is a non-linear method which removes noise from images while keeping the edges unharmed. The neighbouring pixel pattern is called "Window". It slides pixel by pixel inside the whole image. The median is calculated from the window in a numerical order by sorting the entire pixel values. Finally, the replacement has been done with the middle pixel value.

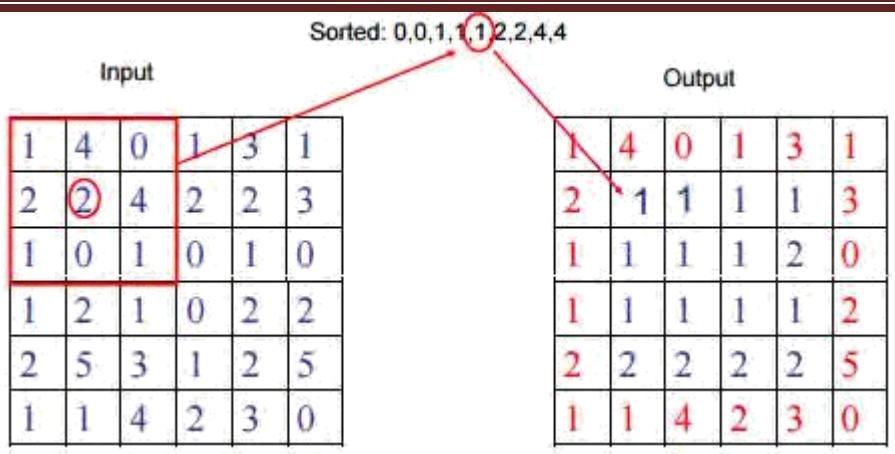


Figure 5.7- Median filtering keeping border values unchanged to preserve edges

5.2.2.4. Adaptive Filter

This is considered as a better option than the previous methods discussed as it is consisted of all their attributes. It requires a filter called Wiener Filter. This method is better than linear filtering because of being selective in nature. Though it has good amount of accuracy, it requires more computation time. Wiener2 function is responsible for primary computations and implementation of the filter.

5.2.2.5. Histogram Equalization

Histogram equalization is a method where intensity of the images is stretched out so that the contrast is adjusted. The intensities are stretched across so that a better pre-processing has been done. Different image processing techniques require Histogram Equalization. We also have used this method as pre-processing technique which is described in work methodologies.

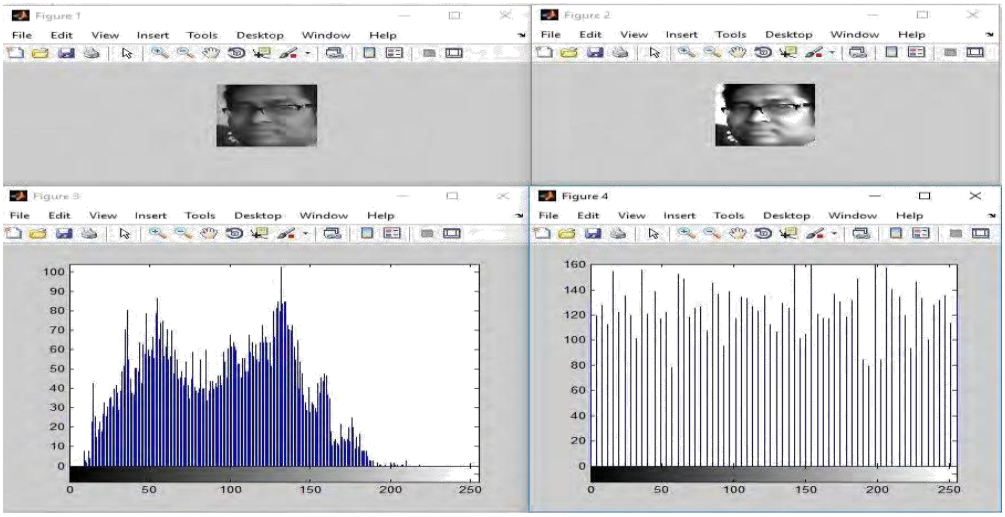


Figure 5.8- Before and after applying Histogram Equalization

5.2.3. Classification

5.2.3.1. Gender Detection Classifier

The Gender Detection Classifier stores pictures for classification purpose and makes classification supported similarity live. Subsequently, Manhattan Distance is calculated between every purpose of the image to the coaching pictures by the classifier. The increment of hit or miss score depends on whether or not the common distance reaches above or below predefined threshold value. Resultant male or female depends on the hit or miss score. If hit score is bigger for specific labels (Male, Female), then the image ought to be thought of as that label.

5.2.3.2. Local Binary Pattern

This method checks points around a central point and tests the surrounding points if they are greater or less than the centre point.

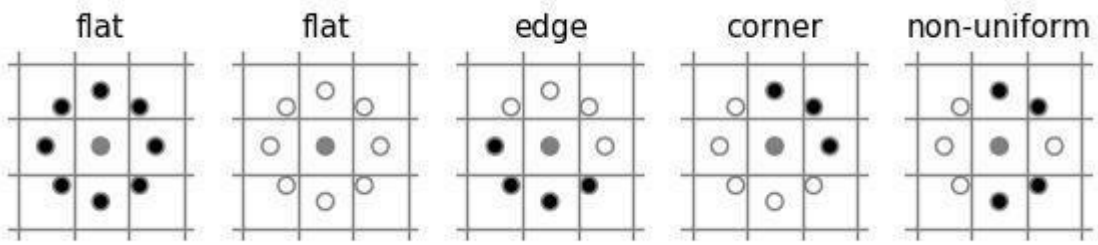


Figure 5.9- Surrounding pixel intensity compared to centre pixel

Points around the central point is checked and tested whether the points are greater or less. We can clearly view the above figure in order to clearly demonstrate this. Pixels which are represented by black or white dots are considered as less or more intense respectively compared to the central pixel. In case of being the surrounding pixels are all black or white, then image region is considered as flat and featureless. Corners or edges are the uniform patterns which are continuous groups of black or white pixels. On the other hand, non-uniform patterns are generated if pixels change to and fro between black and white pixels.

5.2.3.3. Deformable Spatial Pyramid

A Deformable Spatial Pyramid popular for the calculation of dense pixel correspondences. It is a pyramid graph model which matches uniformity at various spatial degrees. Pixel level matching is improved by this method's regularization. Strict rigidity of the previous traditional pyramids has been overcome by the deformable aspect of the model.

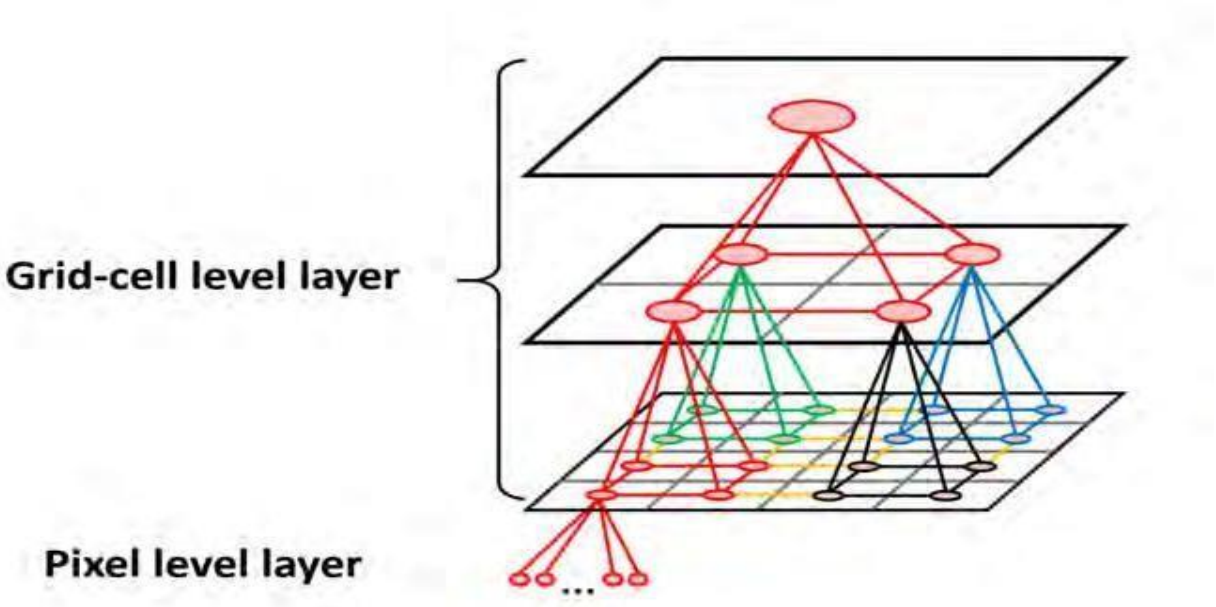


Figure 5.10- The Graph Structure of DSP model

The entire image is divided into four rectangular grid cells. Further it is divided until it reaches to a specific pyramid level number. Finally, a pixel layer is added so that finest pixels have a width of one pixel. The pyramid is represented with a graph. Fixed scale or multi scale matching has been done on the basis of objective. Since the computation time is vast, effective computation is needed.

5.3. Face Recognition

These following problem scope for this project has arrived at after reviewing the literature on face detection & face recognition, & determining possible real-world situations where these systems would be of use. These are following system(s) requirements were identified

1 The system to detect frontal view faces in static images.
2 The system is going to recognize a given frontal view face.
3 Only expressionless and frontal view faces will be presented to face detection & recognition.
4 All the implemented systems must display high degree of lighting in variance.
5 All the systems must have near real-time performance.
6 Both the fully automated & manual face detection must be supported.

7 Frontal view of the face recognition will be realised using only a single known image.

Unfortunately, with these we may specify constricting conditions to our problem domains, it may not be possible strictly adhere to those conditions when implementing a system in the real world.

5.3.1. Fundamental Steps of Image Processing

Fundamental steps in the image processing are: -

1. Image acquisition: To acquire the digital image
2. Image pre-processing: To improve the image in a way that increases the chance for success for the other processes.
3. Image segmentation: To partition an input image into its constituent parts of the objects.
4. Image representation: To convert the input data to the suitable form for the computer processing.
5. Image Description: To extract the features that results in some quantitative information of interest of feature that are basics for differentiating one class of objects from the another.
6. Image Recognition: To Assign a label to the object based on the information provided by its description.

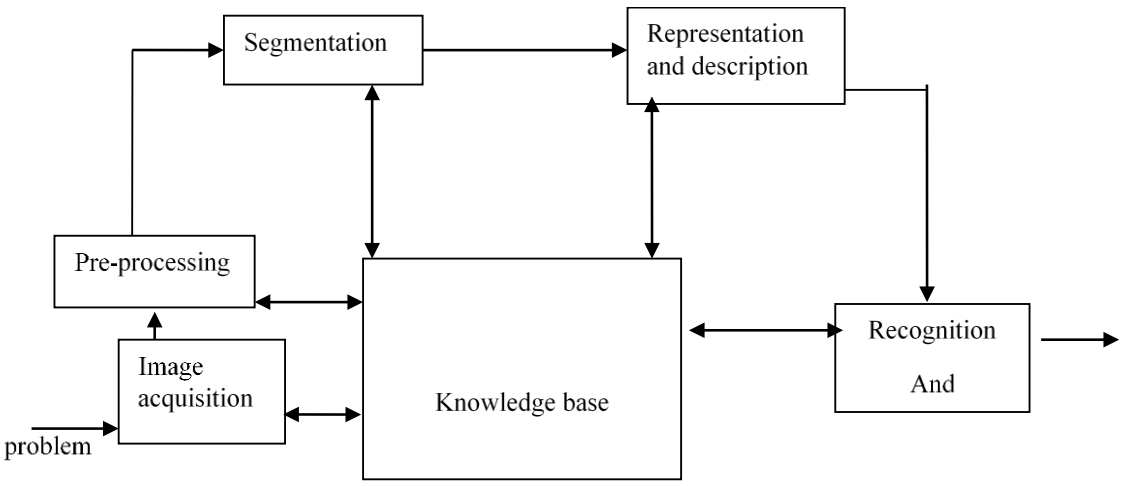


Figure 5.11– Fundamental Steps in Digital Image Processing

5.3.2. Elements of Digital Image Processing Systems

A digital image is processing system contains the following block as shown in the figure

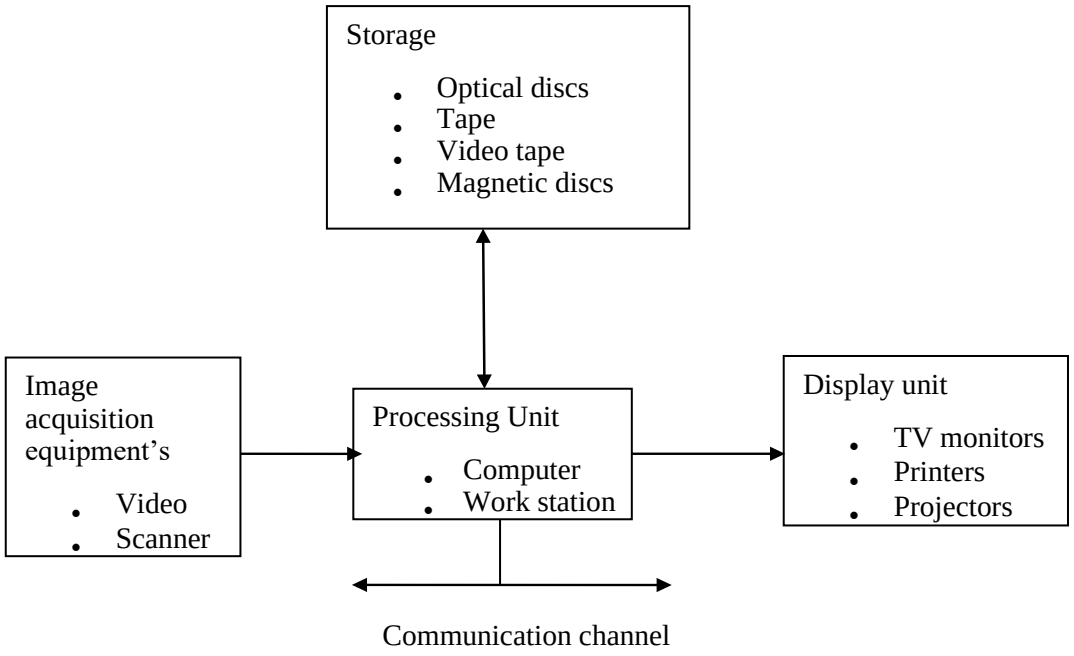


Fig.5.12-Elements of digital image processing systems

The basic operation performed in the digital image processing system which include: -

1. Acquisition
2. Storage
3. Processing
4. Communication
5. Display

Over the previous couple of decades several techniques are planned for face recognition. Much of the techniques planned throughout the first stages of computer vision can't be considered successful, however most of the recent approaches to the face recognition are worthy. According to the analysis by Brunelli and Poggio (1993) all approaches to human face recognition may be divided into 2 strategies:

1. Geometrical features and
2. Template matching.

5.3.3. Face Recognition using Geometrical Features

This technique involves computation of set of geometrical features like nose dimension and length, mouth position and chin form, etc. from the image of the face we would like to acknowledge. This set of features is then matched with the feature of known people. An acceptable metric like Euclidian distance (finding the nearest vector) are often wont to realize the nearest match. Most pioneering add face recognition was done using geometric features (Kanade, 1973), though crop et al. (1987) did comparatively recent add this space.

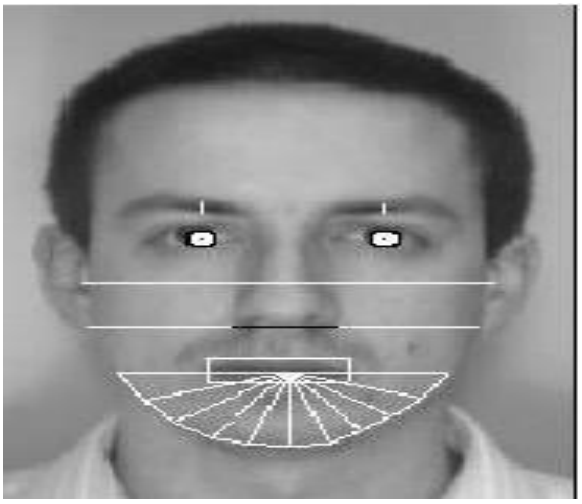


Figure 5.13- Geometrical features (white) which could be used for face recognition

Advantage of using geometrical features as a basis for the face recognition is that in which recognition is possible even at a very low resolutions and with the noisy images (images with much disorderly pixel intensities). The face cannot be viewed in details and its overall geometrical configuration can be extracted for the face recognition. This technique's main disadvantage is that in this automated extraction of the facial geometrical features is very much hard. Automated geometrical feature are extraction based recognition which is also very sensitive to the scaling & rotation of the face in the image plane (Brunelli and Poggio, 1993). This is apparent when we closely examine Kanade's (1973) results where he reported the recognition rate of between 45-75 % with a database of 20 people only.

5.3.4. Face Recognition using Template Matching

This is a very similar the template matching technique which is used in face detection, except here we are not going to try to classify an image as a 'face' or 'non-face' but we are trying to recognize a face.

Figure 5.14- Face recognition using template matching

In this whole face, eyes, nose and mouth regions which can be used in a template matching strategy. The basis of this template matching strategy is to get whole facial regions (matrix of pixels) & compare those with the stored image of the known individual. Once the Euclidean distance can be used to find closest match. Than the simple technique of comparing grey-scale intensity values for face recognition can be used which is by Baron (1981). However, these were far more sophisticated methods of templates matching for the face recognition. These involves extensive pre-processing & transformation of extracted grey-level intensity value. For example, Turk and Pentland (1991a) used Principal Component Analysis & sometimes known as the Eigen faces approach, to pre-process gray-levels and Wiskott et al. (1997) & used Elastic Graphs encoded using Gabor filters to pre-process the extracted regions there. The investigation of geometrical features versus template matching of the face recognition by Brunelli and Poggio (1993) came to conclusion that a feature based strategy may offer higher recognition speed as well as smaller memory requirements, template based techniques offer superior recognition accuracy.

5.3.5. Outline of Implemented System

This Fully automated face detection of frontal view face was implemented using the deformable template algorithm of relying on the image invariants of human face. This is chosen because the similar neural-network are based on face detection model would have been needed far too much training data for implementing & would have been used the great deal of computing time. The major difficulties in implementing the deformable template based on the technique were the creation of bright & dark intensity sensitive templates which designing an efficient implementation of the detection algorithm.

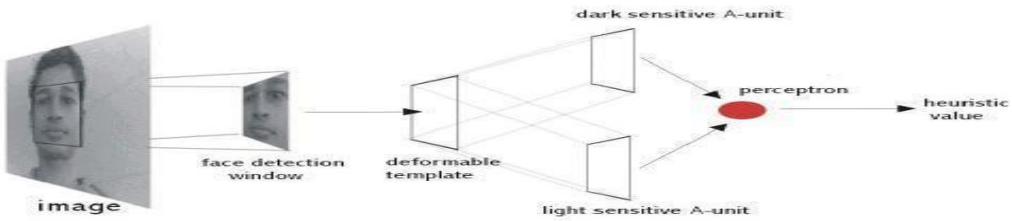


Figure 5.15- Implemented fully Automated frontal view face detection model

A manual face detection system is realised by measuring facial proportions of average face, calculated from the 30 test subjects. To detect the face, the human operator would identify a locations of the subject's eyes in the image & using that proportions of the average face, the system will segment an area from the image.

The template matching based technique is implemented for the face recognition. This was because of its increased recognition of accuracy when we compare to geometrical features based on techniques & the fact that an automated geometrical features are based technique would have required complex features detection pre-processing.

Many possible template matching techniques, Principal Component Analysis is chosen because it has proved to highly robust in pattern recognition tasks & because it was relatively simple to implement. The author also liked to have implemented the technique which is based on Elastic Graphs but could not find the sufficient literature about model to implement these system during the limited time available for the project.

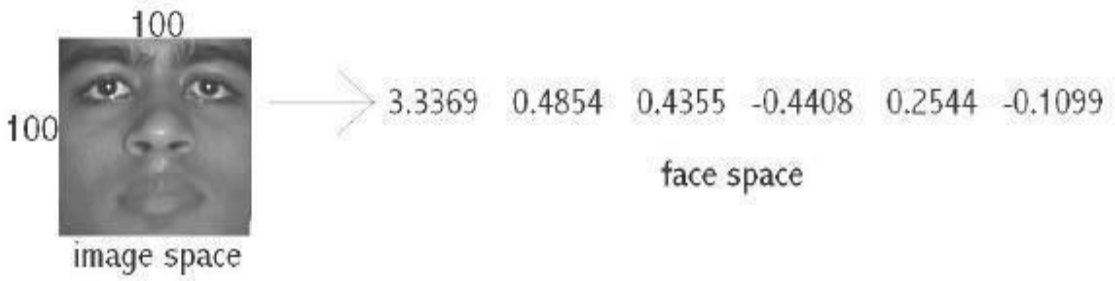


Figure 5.16:- Principal Component Analysis transform from 'image space' to 'face space'.

Using the Principal Component Analysis, the segmented frontal view of the face image is transformed from what is sometimes called 'image space' to 'face space'. All faces in this face database are transformed into the face space.

Face recognition & detection system is a pattern recognition approach for the personal identification purposes, addition to other biometric approaches like fingerprint recognition, signature, retina & so forth. Face is most common biometric used by the human's applications ranges from the static, mug-shot verification in cluttered background.

5.4. Other Modules

5.4.1. PyQt5

Qt is set of cross-platform C++ libraries that implement high-level APIs for accessing many aspects of modern desktop and mobile systems. These include location and positioning services, multimedia, NFC and Bluetooth connectivity, a Chromium based web browser, as well as traditional UI development.

PyQt5 is a comprehensive set of Python bindings for Qt v5. It is implemented as more than 35 extension modules and enables Python to be used as an alternative application development language to C++ on all supported platforms including iOS and Android.

PyQt5 may also be embedded in C++ based applications to allow users of those applications to configure or enhance the functionality of those applications.

5.4.2. OpenCV

OpenCV-Python is a library of Python bindings designed to solve computer vision problems. Python is a general-purpose programming language started by Guido van Rossum that became very popular very quickly, mainly because of its simplicity and code readability. It enables the programmer to express ideas in fewer lines of code without reducing readability.

Compared to languages like C/C++, Python is slower. That said, Python can be easily extended with C/C++, which allows us to write computationally intensive code in C/C++ and create Python wrappers that can be used as Python modules. This gives us two advantages: first, the code is as fast as the original C/C++ code (since it is the actual C++ code working in background) and second, it easier to code in Python than C/C++. OpenCV-Python is a Python wrapper for the original OpenCV C++ implementation.

OpenCV-Python makes use of Numpy, which is a highly optimized library for numerical operations with a MATLAB-style syntax. All the OpenCV array structures are converted to and from Numpy arrays. This also makes it easier to integrate with other libraries that use Numpy such as SciPy and Matplotlib.

5.4.3. OS

The OS module in Python provides functions for interacting with the operating system. OS comes under Python's standard utility modules. This module provides a portable way of using operating system dependent functionality. The *os* and *os. path* modules include many functions to interact with the file system.

5.4.4. Web browser

The webbrowser module provides a high-level interface to allow displaying Web-based documents to users. Under most circumstances, simply calling the open() function from this module will do the right thing.

The script webbrowser can be used as a command-line interface for the module. It accepts a URL as the argument. It accepts the following optional parameters: -n opens the URL in a new browser window, if possible; -t opens the URL in a new browser page ("tab"). The options are, naturally, mutually exclusive.

5.4.5. Psutil

Psutil is a Python cross-platform library used to access system details and process utilities. It is used to keep track of various resources utilization in the system. Usage of resources like CPU, memory, disks, network, sensors can be monitored. Hence, this library is used for system monitoring, profiling, limiting process resources and the management of running processes. It is supported in Python versions 2.6, 2.7 and 3.4+.

5.4.6. Wolfram|Alpha

The Wolfram|Alpha Webservice API provides a web-based API allowing the computational and presentation capabilities of Wolfram|Alpha to be integrated into web, mobile, desktop, and enterprise applications. Wolfram Alpha is an API which can compute expert-level answers using Wolfram's algorithms, knowledgebase and AI technology. It is made possible by the Wolfram Language.

5.4.7. Pyjokes

Python supports creation of random jokes using one of its libraries. Let us explore it a little more, Pyjokes is a python library that is used to create one-line jokes for programmers. Informally, it can also be referred as a fun python library which is pretty simple to use.

5.4.8. PyPDF2

PyPDF2 is a python library built as a PDF toolkit. It is capable of:

- Extracting document information (title, author, …)
- Splitting documents page by page
- Merging documents page by page
- Cropping pages
- Merging multiple pages into a single page
- Encrypting and decrypting PDF files

  and more!

5.4.9. Requests

Requests library is one of the integral part of Python for making HTTP requests to a specified URL. Whether it be REST APIs or Web Scrapping, requests is must to be learned for proceeding further with these technologies. When one makes a request to a URI, it returns a response. Python requests provides inbuilt functionalities for managing both the request and response.

5.4.10. Socket

Socket programming is a way of connecting two nodes on a network to communicate with each other. One socket(node) listens on a particular port at an IP, while other socket reaches out to the other to form a connection. Server forms the listener socket while client reaches out to the server.

They are the real backbones behind web browsing. In simpler terms there is a server and a client.

5.4.11. Pyttsx 3

Pyttsx3 is a text-to-speech conversion library in Python. Unlike alternative libraries, it works offline and is compatible with both Python 2 and 3. An application invokes the pyttsx3.init() factory function to get a reference to a pyttsx3. Engine instance. it is a very easy to use tool which converts the entered text into speech.

The pyttsx3 module supports two voices first is female and the second is male which is provided by “sapi5” for windows.

It supports three TTS engines:

- sapi5 – SAPI5 on Windows
- nsss – NSSpeechSynthesizer on Mac OS X
- espeak – eSpeak on every other platform

5.4.12. PyAutoGUI

PyAutoGUI lets your Python scripts control the mouse and keyboard to automate interactions with other applications. The API is designed to be as simple. PyAutoGUI works on Windows, macOS, and Linux, and runs on Python 2 and 3.

PyAutoGUI has several features:

- Moving the mouse and clicking or typing in the windows of other applications.
- Sending keystrokes to applications (for example, to fill out forms).
- Take screenshots, and given an image (for example, of a button or checkbox), find it on the screen.
- Locate an application’s window, and move, resize, maximize, minimize, or close it (Windows-only, currently)
- Display message boxes for user interaction while your GUI automation script runs.

Chapter 6

TESTING AND RESULTS

6.1 Testing

In this chapter, an overview of testing is provided to verify the correctness and the functionality of the system. Software testing is the process of analyzing a software item to detect the differences between the existing and the required conditions and to evaluate the features of the software item. Software testing is an activity that should be done throughout the development process. Software testing is a task intended to detect defects in software by contrasting a computer program's expected results with its actual results for a given set of inputs.

Test Environment:

The software was tested on the following Environments:

1. Python 3.6

2. Windows environment.

Test Case:

Set of test inputs, execution conditions, and expected results developed for a particular objective, such as to exercise a particular program path or to verify compliance with a specific requirement.

• Features to be tested

• Purpose of testing

• Pass/Fail Criteria

6.1.1 Unit Testing

Unit testing is the testing of individual hardware or software units or groups of related units. Using the unit test plans prepared in the design phase, important control paths are

tested to uncover errors within the boundary of the modules. The interfaces of each of the modules are tested to ensure proper flow of the information into and out of the modules under consideration.

Each unit in this project was thoroughly tested to check if it might fail in any possible situation. This testing was carried out at the completion of each unit. At the end of the unit testing phase, each unit was found to be working satisfactorily in regard to the expected output from the module.

6.1.2 Integration Testing

Integration testing is the testing in which software components, hardware components, or both are combined and tested to evaluate the interaction between them. The various modules are tested for their accuracy and compatibility. The purpose of integration testing is to detect any inconsistencies between the units that are integrated together.

6.1.3 System Testing

System testing is the testing conducted on a complete, integrated system to evaluate the system compliance with its specified requirements. System testing takes, as its input, all of the integrated components that have passes integration testing.

TEST CASE ID	TEST DESCRIPTION	SAMPLE INPUT	EXPECTED RESULT	ACTUAL RESULT	PASS/FAIL
1	Capturing & identifying User's image if registered	Image of the User in proper lighting Conditions	Recognize the user	Recognise the user with ~80% accuracy	PASS

2	Capturing & identifying User's gender	Image of the User in proper lighting Conditions	Recognize the gender of user	Recognise the gender of user with ~90% accuracy	PASS
3	Send face recognised data to the main module	Capture and analyse face	Send the data to the main module for further processing	Module sends data seamlessly and further processing is done	PASS
4	Send gender recognised data to the main module	Capture and analyse face	Send the data to the main module for further processing	Module sends data seamlessly and further processing is done	PASS
5	Send gender recognised data to the main module	Capture and analyse face	Send the data to the main module for further processing	Module sends data seamlessly and further processing is done	PASS
6	Getting calendar data from google	An authorized user's image and data	Receive the Google's calendar data	Received the Google's calendar data	PASS
7	Calculate Mathematical equations	An equation with speech	Should calculate any valid	Calculation of valid mathematical question	PASS

			mathematical equations		
8	Gather weather data for the interface	An openWeather API	Weather data to be received	Weather data Received	PASS
9	System testing in various versions of OS	Execute program in windows 8 and above	Performance to be better in Windows	Same as expected, performance better for windows	PASS
10	Sending message for unauthorized access	Unauthorized user & Voice command, i.e. do I have plans today?	Message to be received	Message is received	PASS

Table 6.1. Test Cases

RESULTS:-

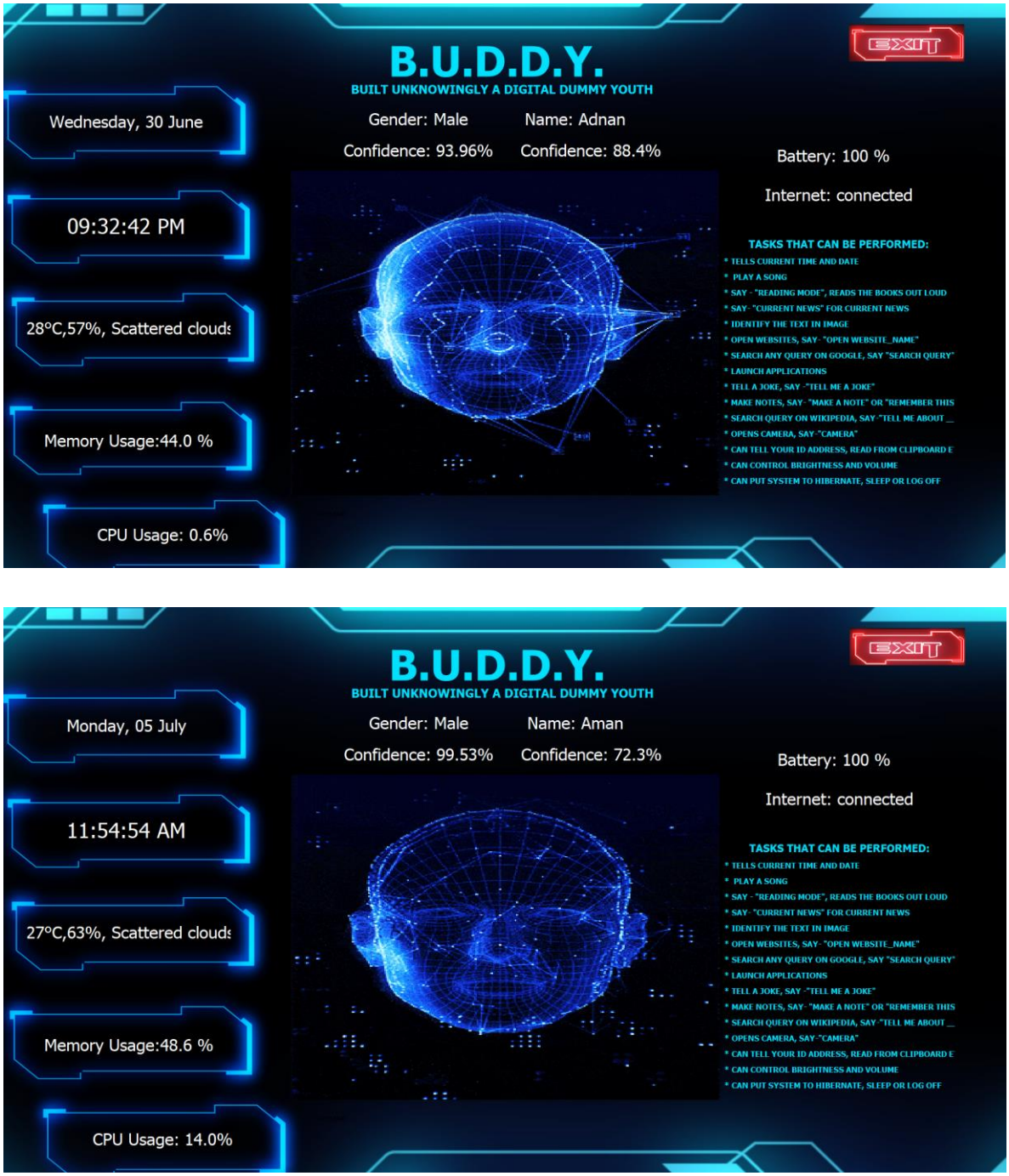


Figure 6.1- Detection of two different people by assistant

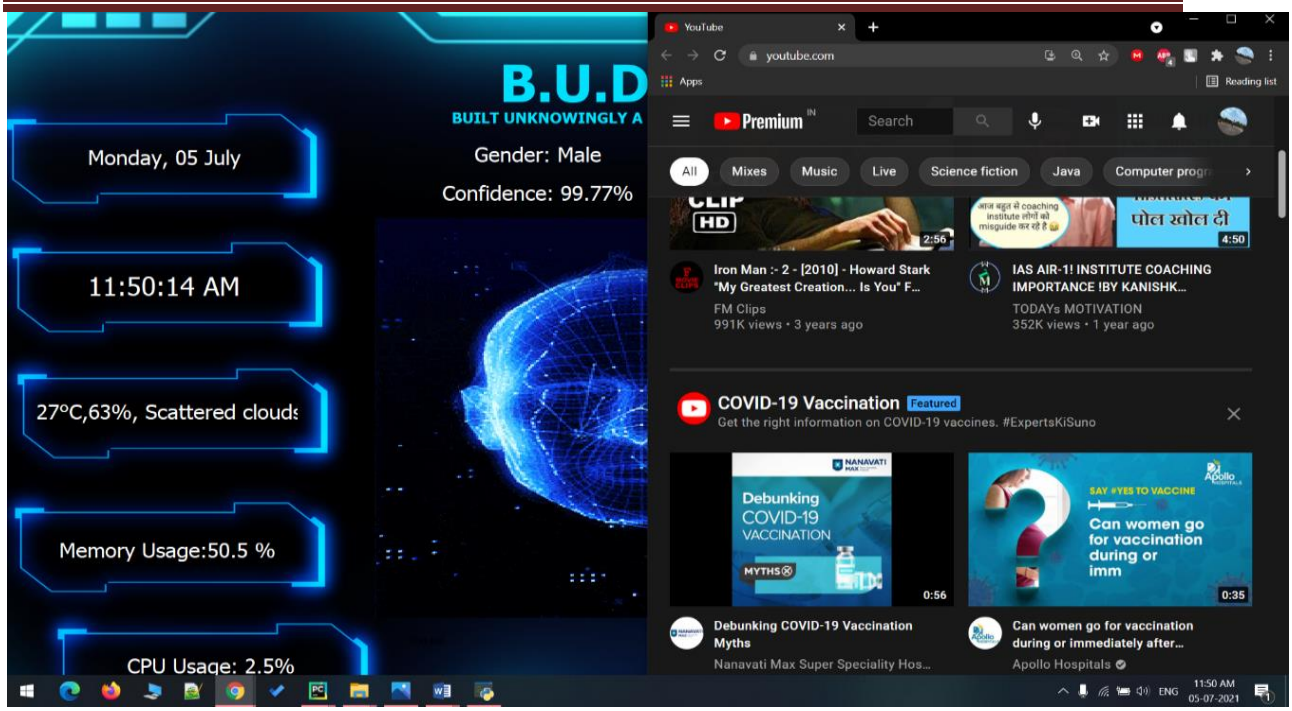


Figure 6.2- Opening websites (e.g., YouTube)

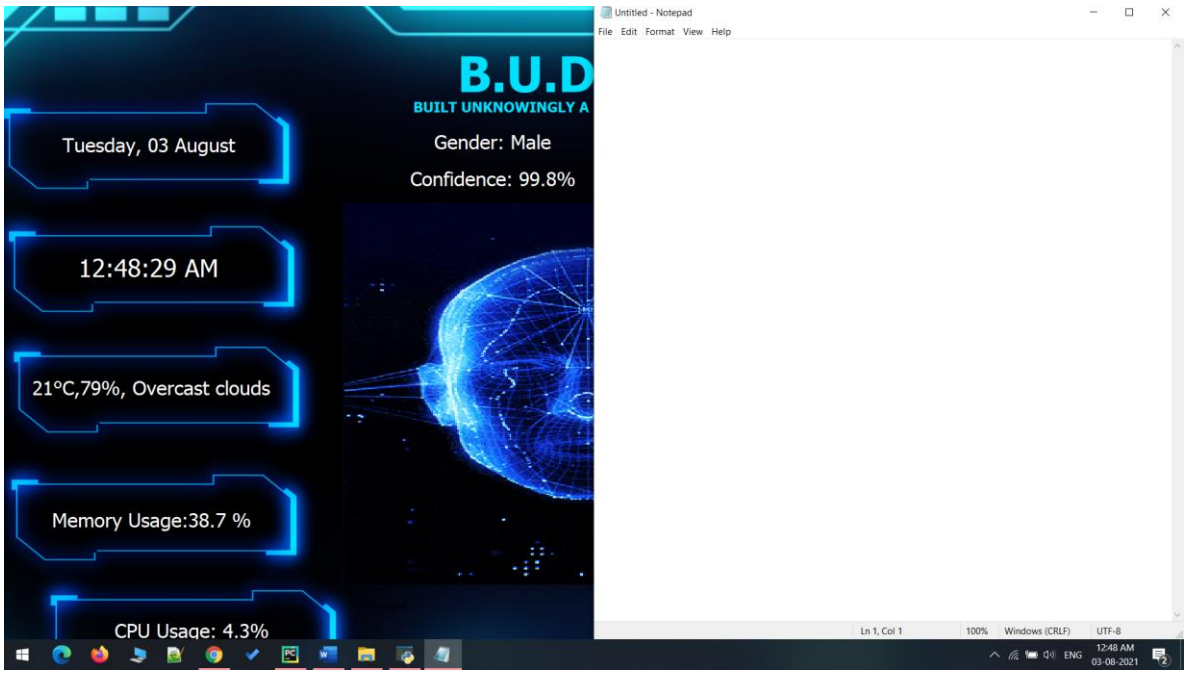


Figure 6.3- Launching Applications (e.g., Notepad**)**

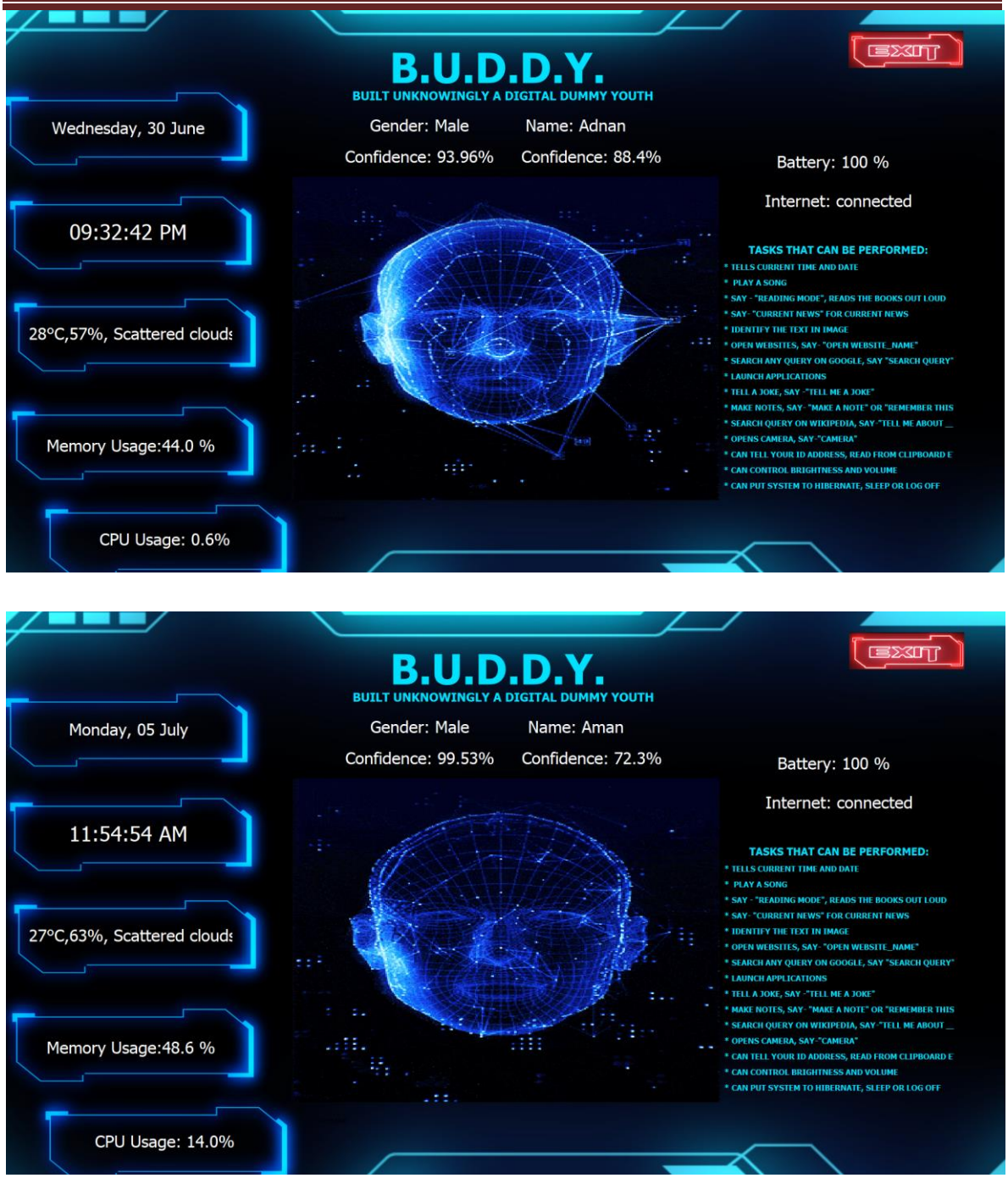


Figure 6.4- Launching other application.

Chapter 7

CONCLUSION AND FUTURE ENHANCEMENTS

The personal voice assistant system given during this paper is extremely basic system with few options but the extra and advance feature could also be introduced as future work of this project, during this paper the look and implementation of a Intelligent Personal Voice help is delineate. The project is made mistreatment offered open supply code modules with visual studio code community backing which might accommodate any updates in future. The standard approach utilized in this project makes it a lot of versatile and simple to integrate further modules and options while not heavy the present system functionaries. It not solely works on human commands however conjointly it's designed for offer responses to the user on the idea of question being asked or the words spoken by the user like gap tasks and operations. This Intelligent Voice Assistant has a massive and limitless scope within the future. Like Siri, Google currently and Cortana most well-liked personal voice assistants. The project can simply able to integrate with devices close to future for a Connected Home mistreatment web of Things, voice command system and pc vision.

REFERENCES

[1] Speech Recognition Using Deep Neural Networks: A Systematic Review February 2019IEEE Access PP(99):1-1

[2] (IJCSE) International Journal on Computer Science and Engineering, Vol. 02, No.01S, 2010, 14-23 14 Feature Extraction based Face Recognition, Gender and Age Classification

[3] A Study on Face Recognition Techniques with Age and Gender Classification

[4] International Journal of Engineering Trends and Applications (IJETA) – Volume 4 Issue 2, Mar-Apr 2017

[5] Automated Speech Recognition System – A Literature Review Manjutha M [1] , Gracy J [2] , Dr. P Subashini [3] , Dr. M Krishnaveni [4]

[6] Aditya Sinha, Gargi Garg, GouravRajwani, Shimona Tayal, "Intelligent Personal Assistant". International Journal of Informative & Futuristic Research, Volute. 4, Issue 8, April 2017.

[7] Emad S. Othman ."Voice Controlled Personal Assistant Using Raspberry Pi". International Journal of Scientific and Engineering Research Volume 8, Issue 11, November-2017.

[8] Bibek Behera, "Chappie - A Semi-Automatic Intelligent Chatbot".

[9] MoustafaElshafei, "Virtual Personal Assistant for Mobile Users".

STUDENT DETAILS

NAME – ADNAN ANJUM

USN – 1SG17CS003

SEM, SEC – 8TH , A

PHONE NUMBER – 9162945996

EMAIL ID – adnananjum12@gmail.com

NAME – AMAN PANDEY

USN – 1SG17CS008

SEM, SEC – 8TH, A

PHONE NUMBER – 9106922473

EMAIL ID – amanpandeyfeb2000@gmail.com