



(Established under Galgotias University Uttar Pradesh Act No. 14 of 2011)

DESKTOP ASSISTANT PROJECT BASED ON VOICE RECOGNITION AND FACE DETECTION

A Report for the Evaluation of Project 2

Submitted by

VISHVAKETAN GAUR

(1713104108 / 17SCSE104110)

*in partial fulfillment for the award of the degree
of*

Bachelor of Computer Application

IN

Computer Science and Engineering

SCHOOL OF COMPUTING SCIENCE AND ENGINEERING

**Under the Supervision of
Mr. Soumya R. Jena
Assistant Professor**

APRIL / MAY- 2020

DECLARATION

I **VISHVAKETAN GAUR** hereby declare that the Project work, which is presented in the report, entitled “**DESKTOP ASSISTANT BASED ON VOICE RECOGNITION AND FACE DETECTION** ” is duly prepared by me to be submitted to the department in partial fulfilment for the award of the degree of Bachelor of Computer Applications for the academic year 2019-2020.

VISHVAKETAN GAUR

Date:

Place:

DEPARTMENT OF BACHELOR OF COMPUTER APPLICATION
GALGOTIAS UNIVERSITY, GREATER NOIDA, 201310

BONA FIDE CERTIFICATE

This is to certify that this dissertation entitled “**DESKTOP ASSISTANT BASED ON VOICE RECOGNITION AND FACE DETECTION**” who carried out the work under my supervision. This is to further certify to the best of my knowledge that this Project Report has not been carried out earlier in this institute and university.

Signature
Mr. Soumya R Jena

Signature
Mr. U Samson

Signature
Mr. Nagarajan M

Certify that the above mention Project Report has been duly carried out as per the norms of the college and the statutes of the university.

Signature
Dr. Munish Sabharwal

ACKNOWLEDGEMENT

Every Project big or small is successful largely due to the effort of a number of wonderful people who have always given their valuable advice or lent a helping hand. I sincerely appreciate the inspiration, support and guidance of all those people who have been instrumental in making this project a success.

I wish to express sense of gratitude to my guide Mr. Saumya Ranjan, Mr.U Samson(Project Coordinator), Mr. Nagarajan M(Panel In Charge) in Bachelor of Computer Application Department, Galgotias University, Greater Noida, to give me guidance at every moment during my entire report and giving valuable suggestions. They gives me unfailing inspirations and whole hearted co-opertions in carried out my Project work. Their continuous encouragement at each work through our grateful acknowledged.

I am also very grateful to my classmates, for their huge co-operation and valuable suggestion from time to time during my entire Project work. I also extend my gratitude to all the members of Department , without theirs support at various stages this report will not be materialized.

VISHVAKETAN GAUR

Table of Contents

S.NO.	TITLE	PAGE NO.
1.	Abstract	i
2.	Introduction	1-2
3.	Related Work	3-4
4.	System Design	5
5.	Implementation or architecture diagrams	6-7
6.	Proposed Method	8-10
7.	Hardware and Software Requirements	11
8.	Output of the Project	12-13
9.	Conclusion and Future Enhancement	14-15
10.	References	16

FIGURE INDEX

S.No	Title	Page No.
1.	Face Detection System Architecture	6
2.	Flow-Chart of Voice Recognition	7
3.	Command: Open Google	12
4.	Command: Open RF Campus	12
5.	Command: Open Music System	13
6.	Face Detection	13

ABSTRACT

In very recent future all the electronic devices will be worked by utilizing the remote helper which is definitely not hard to get to yet it needs weakness. This structure affirmed the clients to get to the framework by the voice orders. User can request to the assistant that anything can be done by the framework, for example Music, Open Specified Application, Open Tabs, Open Websites and so forth. Voice associates are programming specialists that can decipher human discourse and react through orchestrated voices. Clients can pose their associate's inquiries, control home gadgets and media playback by means of voice, and oversee other essential errands, for example, email, daily agendas, and schedules with verbal orders.

CHAPTER-1

INTRODUCTION

1.1 INTRODUCTION

Voice assistants are defined as the software agents which interpret or converts human speech and it responds through synchronized voices, i.e. Siri (Apple Inc.), Alexa (Amazon), Cortana (Microsoft), so forth these are most popular voice assistants.

As the technology are getting advancer day-by-day, Though We can see that the futuristic dream of talking to the computer is now comes true As, From the day the computer invented, People wanted to talk with the computer machine, It could be possible with the technology development by the dedicated and devoted computer scientists.

Several products deployed in the last few years which brings out the in-expensive use of the voice assistants to our day-to-day life, With regard to the time more features and platforms are being added and gets updated from time to time.

Intelligent Personal Assistant, Automated Personal Assistant or Automated Virtual Personal Assistant can perform tasks, or services on behalf of an individual based on a combination of user input, and location awareness. It has the ability to access information from a variety of online sources such as weather conditions, traffic congestion, news, stock prices, user schedules, and retail prices.

Intelligent Personal Assistants features include calendar and Meeting Reminder, Automation, Natural Conversation, Recommending, Smarter Learning and Integration.

Voice Assistants can make calls, send text messages, look things up online, provide directions, open apps, set appointments on our calendars, and initiate or complete many other tasks. With the addition of separate apps on the phone, our voice can be a type of remote control for our lives.

1.2 MODULES AND DESCRIPTION

- **SPEECH TO TEXT:** In this module or layer a speech is to be converted into the text, Which can be understandable by the system through the installed libraries.
- **TEXT ANALYSING:** It converts the text for the system, it analyze and then make it into the readable form. Though, Computer understands the command, so virtual assistants like Siri, Cortona so forth converts the text to computer commands.
- **INTERPRET COMMANDS:** In this layer, mapped PC sends requests to the server through the web. At the same time, as the discourse assessed locally. A nearby recognizer speaks with the server to decide whether the order will be ideal to deal with locally or not. For example Play Music, Restaurant Reservation, Movie Review and so forth.

CHAPTER-2

RELATED WORK

2.1 RELATED WORK

This system has a long history with a few rushes of significant advancements. It acknowledges for transcription, search, what's more, voice orders has become a standard component on cell phones and wearable gadgets. [1] It is accomplished for utilizing a CTC based LSTM an acoustic model, It predicts setting free telephones so forth, It shrink to a 1/10 of its unique size utilizing a mixture of SVD-based pressure and quantization. It restricts the networks and on-the-fly language model re-scoring to accomplish constant execution on present-day cell phones. In [6] we likewise coordinate an inquiry parsing module among ASR and Search for multiple reasons. Also it [3] contains the methods for the proper implementation of computer generated voice search it administrates and proposed the versatile clients getting to the administrations over a scope of compact gadgets. Voice recognition is executed as 2-phase recognizing strategy where string up-and-comers created by a programmed Automatic Speech Recognition (ASR) framework. So as to distinguish the coordinating passage from a possibly extremely enormous application-explicit DB that gives a genuine case of how extra explicit information sources can be utilized with an ASR framework to encourage voice access to online hunt files.

In [9] author's precision is to prepare bigger acoustic models. There is a nonparametric system, the exact model that abuses rich preparing information to legitimately learn elocution variety. Introducing the exact model with a parametric model performs efficiently, with an overall improvement of 5.2% in WER [2]. [7]There are various methods in which this work could be expanded. In the first place, it reconciles with the acoustic model, preparing is probably going to be a keener appropriations and a more tightly fixed to the information. Secondly, evaluating word articulation co-event includes in semi-managed style (for example through word acknowledgement rather than constrained arrangement) would widen its relevance to a wide scope of various types and errands. The observations from an overlooking literature review in

order to present universal knowledge (theory and concepts) about voice control, digital assistants, fields of use and more (Stufflebeam and Shinkfield, 2007). [4]Computerized database key word looking became used on this review, As the technique is fast and efficient. The 3-Fundamental DB used for the studies changed into Google, Google Scholar and the KTH Publication database DiVA. Google turned into to discover general information as Google Scholar and DiVA were used to locate most research facts.

At the end, it is to be good to change our models to factor with the distinct phenomena that affect the pronunciation (e.g. accent, dialect, recognition errors) [5]. This paper highlights AT&T Speak4it R voice search application which points out the development and advances on automatic speech recognition. Security is the major issue with the voice-assisted devices whereas anyone can access to a voice-activated device. He can ask it questions, collects information about the accounts and services associated with the active device, and ask it to perform tasks which can be hazardous and can cause loss too. It consists of high security risk as these devices may expose the calendar contents, emails, and other personal details. [6] It was reported by a person that, his iPad in his living room would unlock the front door for anyone who stood outside and asked Siri to let them in. Recently Google has upgraded its Assistant soft-ware. The upgraded version includes various features like voice printing, which uniquely identifies each user by voice [8].

CHAPTER-3

SYSTEM DESIGN

3.1 SYSTEM FRAMEWORK

SYSTEM DESIGN:

It is characterized as the way toward defining the segments for example designs, modules, interface, and information for an offered framework to fulfill determined necessities. Framework configuration could be viewed as the use of framework hypothesis to item advancement. It covers with the control of a framework investigation, framework design and framework building.

PRINCIPLES OF SYSTEM DESIGN:

Try not to be a human:

Playing lure switch with a client can cause them to feel that they been tricked, or that they don't see how a framework functions; both are terrible encounters. Maintain a strategic distance from, Indicators or fake postponements to cause the UI to appear to be increasingly human, Interaction between the client and the bot framework plainly marked in a manner that conveys progressively human.

Keep It Simple:

Discussion ought to be limited to specific subjects and follow straight discussions streams and do maintain a strategic distance from muddled stretching ways. It's alright to uncover and clarify confinements. As users will tire of complicated passage of dialogue.

Utilize structured information when conceivable:

Try not to put clients in a circumstance where they have to figure the right mantra required to continue. Custom Soft consoles license a restricted scope of information and can spare a lot of composing. For example as opposed to requesting that the end client type "YES" or "NO".

CHAPTER-4

ARCHITECTURE DIAGRAMS

System Architecture Design divided into two phases i.e. Face Detection and Voice Recognition

4.1 SYSTEM ARCHITECTURE DESIGN

Face Detection:

Create a database in which all the training images are stored and the captured images to be stored in database. The face detection to be completed by two algorithms i.e. Haar Cascade Algorithm and Fisher Face Algorithm. In the Haar Cascade, Calculation depends on Course classifiers which comprises of Haar highlights. The course classifier is interlinked with lots of classifiers which helps to detect the human face it subject to the most noteworthy features like eyes, eyebrows and lips. Haar highlights are distinguished and depends on the solid calculation, in which we assign a pixel force to each and every pixel identified with grayscale values inside the extent of 0 to 255 where 0 for the white and 255 for the dark-shading.

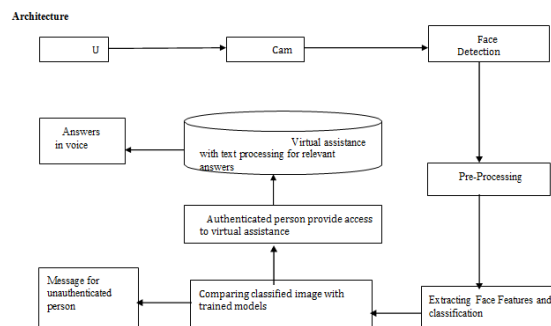


Fig.1 Face Detection System Architecture

Voice Recognition:

Personal Assistant consists of various modules services, but the salient feature of the assistant is that Voice Recognition can function without active internet.

Flow Chart of the Voice Recognition:

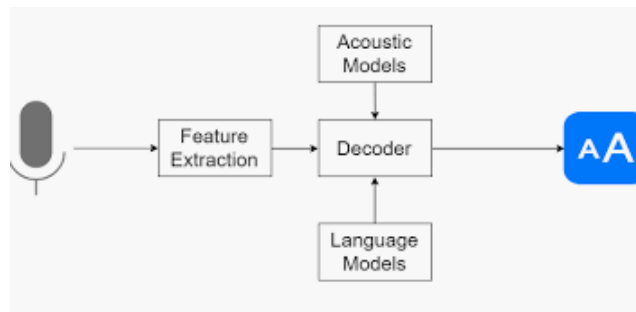


Fig.2 Flow-Chart of Voice Recognition

CHAPTER-5

PROPOSED METHOD

ALGORITHM USED:

Algorithm:

Voice Recognition: Python Speech Recognition Module

This algorithmic program are the modules or the packages for the recognizing the voice.

1. *“sudo pip install SpeechRecognition”*
PyAudio(For Linux Users)
2. *“sudo apt-get install python-pyaudionpython3-pyaudion”*
PyAudio(For Windows User)
3. *“ pip install pyaudio”*

Face Detection:

Haar Cascade Algorithm used for the face detection which follows the several steps:

1. In this algorithm, It depends on the Course classifiers which consist of Haar feature which are stored in Haar record for the identification of face.
2. These classifiers are the group of weak classifiers used to create a strong classifier.
3. This forms a triangle shape which helps to identify the color of the image. i.e. Black or White

Face is recognized using the fisher face algorithm:

This involves in training the images in which are stored datasets. It creates the different records one for pictures/images and other ones images with their comparing names. It Encircle the picture and their id for every single sub-directory in the dataset. Then create a numpy array

common for both the lists. Open- Cv helps to train models for images with respective id using function.

1. It tells about a class-explicit change grid.
2. It analyze discriminantly the facial features to compare between the persons.
3. It heavily depends on the input data as well.
4. It also allows for a reconstruction of the projected image.

Implementation:

Datetime: In python this module supplies classes to work with date and time. And helps to deal with the dates, times, and time intervals.

Open-Cv: This module sorts the image and analysis the video like face detection, license plate reading, optical character recognition, and so forth.

Httpplib: This module helps for the implementation of the client side of the Hyper Text Protocols(HTTP and HTTPS). Generally, It can't be accessed directly but uses URLLIB which uses to handle URLs.

Urllib2: This module defines the functions and helps to open the URLs but mostly HTTP i.e. Authentication, Redirections, Cookies and more.

Json: It's a JavaScript Object Notation., It's a lightweight data interchange format derived by **JS**. It exposes an API familiar to users of the standard library marshal and pickle modules.

Subprocess: This module allows spawning new processes, Which connect to their onput/output/errors and obtain their return codes. This module used to replace several older functions and modules as well.

Cv2: This python method loads an image from the specified file and read it. It consists of three flags which are as follows i.e. cv2.IMREAD_COLOR, cv2.IMREAD_GRAYSCALE, cv2._UNHANGED

Numpy: A general purpose array processing package which provides a multidimensional array object, and tools for working with these arrays. It's a basic package for scientific computing with the python.

OS: This module provides a way of using operating system dependent functionality. OS module allows us to interface with the underlying operating system that on python runs.

ADVANTAGES

- It helps to save the time by assigning the tasks, also can assign the tasks repeatedly.
- Helps to deal with daily routine, For Example: By assigning the task for ringing the alarm at certain time, also can schedule the appointment and so forth.
- It makes the work faster easier and in the efficient manner.

DISADVANTAGES

- Language issue may arise.
- Strong risk of downtime
- May sometime cause network glitches

APPLICATIONS

- Desktop Assistant can be used int the day-to-day routine which can save the time.
- This assistant can likewise be installed on the site of the corporate areas, lodging and the executives businesses for questionnaires and so on.

CHAPTER-6

HARDWARE AND SOFTWARE REQUIREMENTS

HARDWARE REQUIREMENTS:

- i3 Processor(Or More) required.
- 4GB(Or More) required.
- Monitor.

SOFTWARE REQUIREMENTS:

- Windows7(Or Higher Config.) required.
- Kernel Version 3.0.16 or Higher
- Active Internet Connection.

CHAPTER-7

OUTPUT OF THE PROJECT

Here is the output result of the following desktop assistant.

Voice Recognition:

In this phase an user speaks to the desktop assistant, It recognize the voice command encode it and answers to the user.

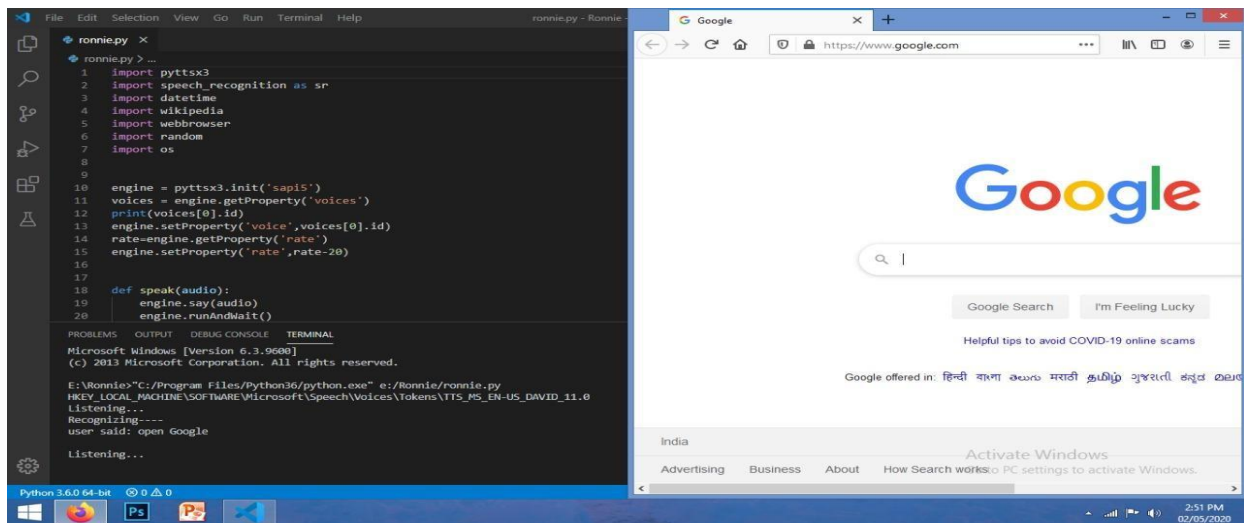


Fig.3 Desktop Assistant (Voice Recognition)
(Command: open Google)

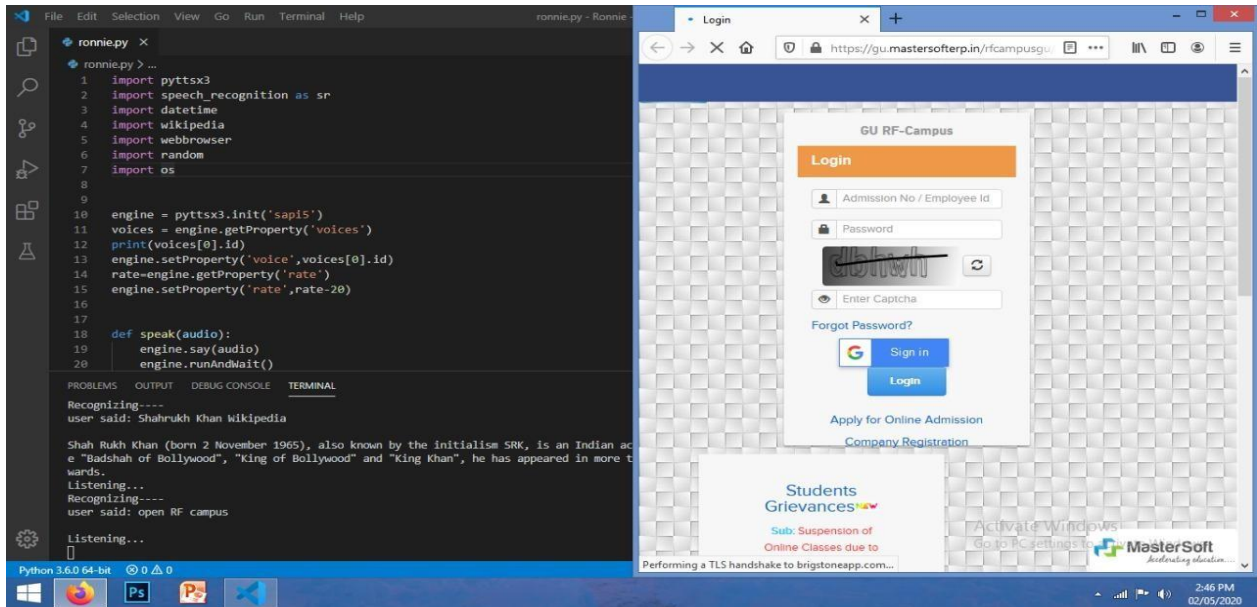


Fig.4 (Command: Open RF Campus)

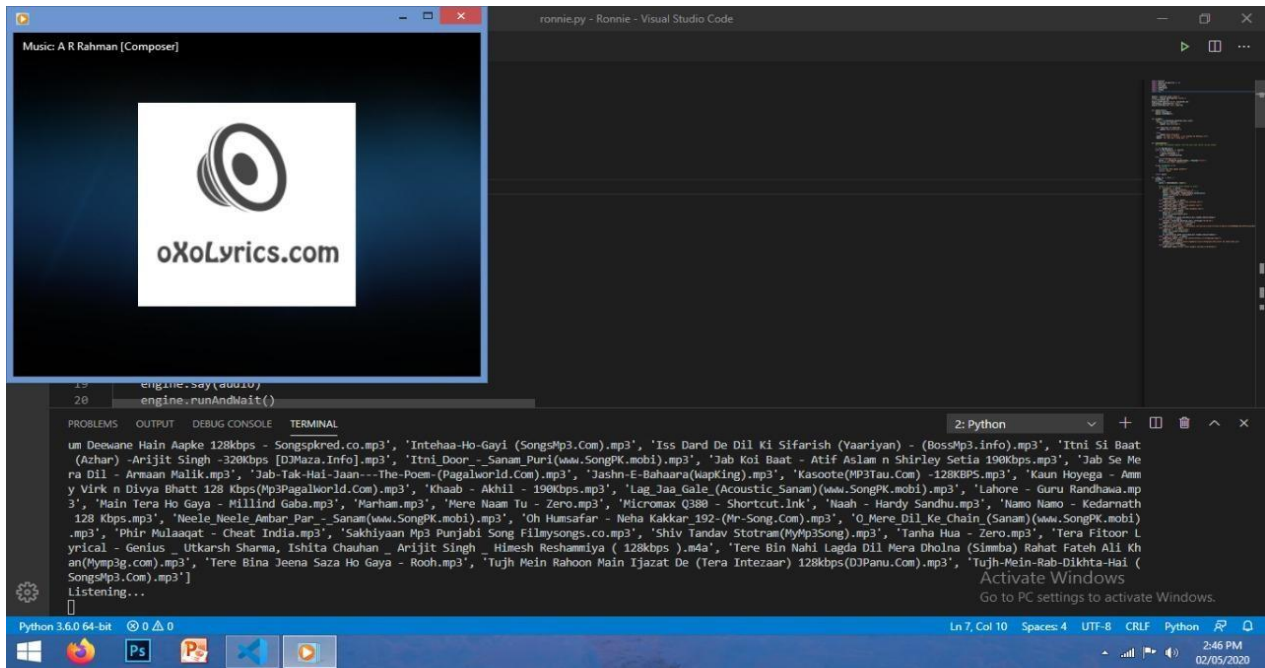


Fig.5 (Command: Open Music System)

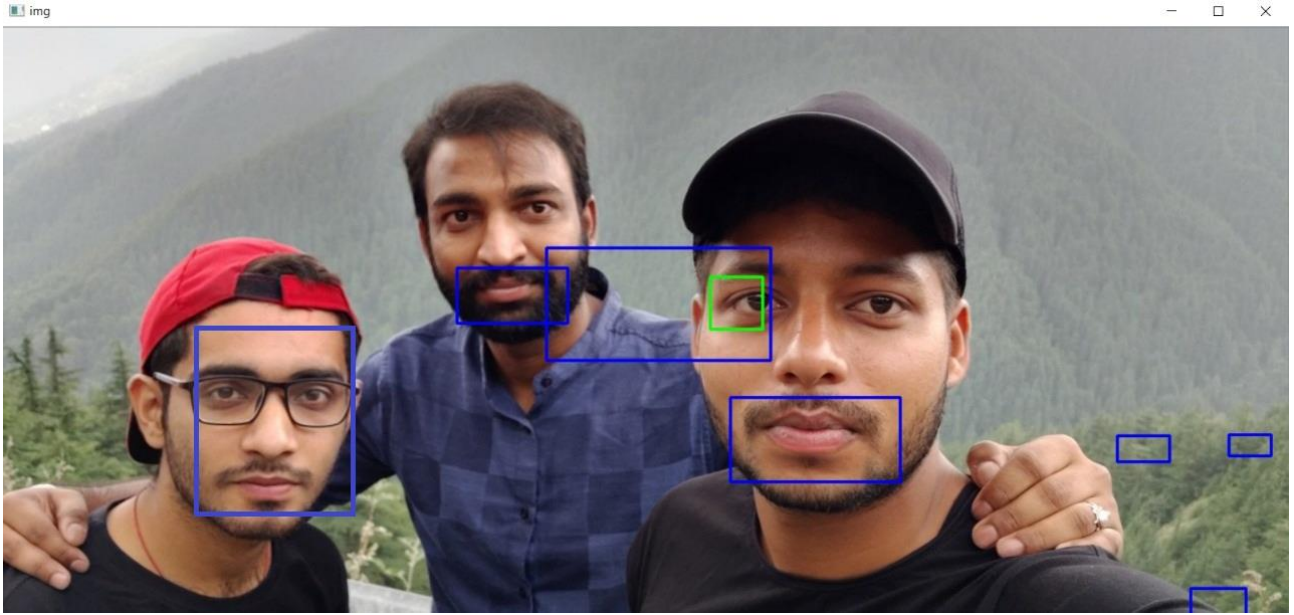


Fig.6 Face Recognition

CHAPTER-8

FUTURE PROSPECTS OF THE PROJECT

8. FUTURE PROSPECTS

How rapidly the time changed? If we look back around twenty year ago, Voice Recognition was in its infant stage. As when the computer system came into the existence, It was the dream to fully fledged interaction with the computer machine. Now, we can eventually talk, ask and as well order to do the assigned task.

This technology advancement is taking the world to the next level. If we think about the future competence of the voice recognition as well as the face detection, It can help the security agencies which may help to verify the details of the criminal and so forth. If we compare, two decades ago our words may very carried as we could have imagined.

CONCLUSION

Desktop Assistant has various functions as like of mobile phone like managing various applications on the voice commands. It helps to access the system hands-free and to get rid of typing chaos.

By the use of facial recognition an individual can access the system, with the help of the face detection helps to secure the data, as no other person can access the system. It strategies the machine learning and help the user to access securely.

REFERENCES

- [1] Xin Lei, Andrew Senior, Alexander Gruenstein and Jeffrey Sorensen “Accurate and Compact Large Vocabulary Speech Recognition on Mobile Devices,” in INTERSPEECH. 2013, pp. 662–665, ISCA.
- [2] Alex Graves, Santiago Fernández, Faustino Gomez, and Jürgen Schmidhuber “Connectionist Temporal Classification: Labelling Unsegmented Sequence Data with Recurrent Neural Networks,” Proceedings of the 23rd International Conference on Machine Learning, Pittsburgh, PA, 2006.
- [3] Brian Kingsbury, “Lattice-based Optimization of Sequence Classification Criteria for Neural-Network Acoustic Modeling”, 2009 IEEE International Conference on Acoustics, Speech and Signal Processing, Taipei, Taiwan.
- [4] Daniel L. Stufflebeam, and Anthony J. Shinkfield, “Evaluation Theory, Models, and Applications”, John Wiley & Sons, 2007.
- [5] David Rybach, Michael Riley, and Chris Alberti, “Direct Construction of Compact Context-dependency Transducers from Data,” Proc. of INTERSPEECH, pp. 218–221, 2010.
- [6] K. Beulen ; and H. Ney “Automatic Question Generation for Decision Tree Based State Tying,” Proceedings of the 1998 IEEE International Conference on Acoustics, Speech and Signal Processing, ICASSP '98.
- [7] Patrick Nguyen, Georg Heigold, Geoffrey Zweig, “Speech Recognition with Flat Direct Models,” IEEE Journal of Selected Topics in Signal Processing, Volume: 4 , Issue: 6 , Dec. 2010.
- [8] Klaus Beulen, Hermann Ney, “Automatic Question Generation for Decision Tree Based State Tying,” Proceedings of the IEEE International Conference on Acoustics, Speech and Signal Processing, ICASSP '98.

