



INTERNATIONAL JOURNAL OF ADVANCE RESEARCH, IDEAS AND INNOVATIONS IN TECHNOLOGY

ISSN: 2454-132X

Impact factor: 6.078

(Volume 6, Issue 4)

Available online at: www.ijariit.com

Controlling web browser via different hand gestures and speech recognition

Rithika V. Jadhav

rithikajadhav1407@gmail.com

Usha Mittal Institute of Technology,
Mumbai, Maharashtra

Prajakta S. Bait

prajakta.bait1@gmail.com

Usha Mittal Institute of Technology,
Mumbai, Maharashtra

Sumedh Pundkar

sumedh.pundkar@umit.sndt.ac.in

Usha Mittal Institute of Technology,
Mumbai, Maharashtra

ABSTRACT

Gestures and speech are major forms of human communication. Primary goal of gesture recognition is to create a system that can identify specific human gestures whereas primary goal of speech recognition is to create a system that can convert speech sound into its corresponding text. In this paper we have implemented a real time gesture recognition system where a user can navigate a computer to a particular website by doing a specific gesture in front of a web camera linked to the computer. We have also implemented a speech recognition system which takes speech as an input from the device microphone, converts it into text and thereafter performs different operations.

Keywords— Gesture, Speech, Human Communication, Gesture Recognition, Speech Recognition

1. INTRODUCTION

Humans use different sensory modes like gesture, speech, facial & body expressions to interact with each other. This goes a long way to smoothen and improve the relationship between them. Similarly, the basic goal of Human-Computer Interaction is to improve the interaction between users and computers by making computers more receptive to user needs. Today interaction between Human and Computer is not just limited to keyboard and mouse interaction, but interacting with the system naturally is becoming even more important in many fields of Human - Computer Interaction. Hands are human organs which are used to manipulate physical objects. Gestures as well as speech can also be used as a tool of communication between computers and humans. For this very reason hands and speech can be used most frequently by human beings to communicate and interact with machines. Human Gesture Recognition is designed to cater to the needs of the fast-growing world and to minimize the hardware caring burden. On the other hand speech recognition technology provides ways in which computer technology can be utilized for tasks which use natural (human) language or speech.

Mouse and Keyboard are the basic input / output to computers and the use of both these devices require the use of hands. Most important and immediate information exchange between man and machine is through visual and aural aid, but this communication is one-sided. Although hands are most

commonly used for day to day physical manipulation related tasks, in some cases they are also used for communication. Hand gestures and speech support us in our daily communication to convey our messages clearly. Hands are most important for mute and deaf people, who depend on their hands and gestures to communicate. So, hand gestures are vital for communication in sign language. If computers had the ability to translate and understand hand gestures, it would be a leap forward in the field of human - computer interaction. The dilemma faced with this is that the images these days are information rich and in-order to achieve this task extensive processing is required.

However, even the Human-Computer has limitations, where a user or owner of a system does not have his hands free to type in anything using a keyboard or where the user is aged or handicapped. The ability to control his device or system in such a case can be done by speech / voice thereby enabling the user or owner to be privileged to use his device more effectively than getting someone to do it for him. Speech recognition or voice recognition technology takes audio input for entering data rather than a keyboard which means speaking into a microphone and then producing the same result as words are typed manually using a keyboard. In the last few years speech recognition technology has been developed to a great extent.

2. PROBLEM STATEMENT

Internet is growing at a very fast pace. The use of web browser is also growing simultaneously. Nowadays, everyone has at least 3-4 most frequently visited websites. What if visiting those websites can be made simpler? Won't it be a wonderful idea? We plan to address the current situation and try to make the use of a web browser easier. Web Browser is a very common software application which everybody uses every day, every hour and every minute. So, why not try hand gestures and speech recognition to control the web browser itself?

By using this technology, the users can use a web browser in a more easy and efficient way and control it simultaneously.

3. LITERATURE SURVEY

3.1 Literature Survey of Hand Gesture

Several methods are proposed for both dynamic and static hand gestures.

- A robust marker for hand gesture recognition system is designed by Aashni Haria in [4] which can efficiently track gestures.
- Viraj Shinde, Tushar Bacchav in [2] proposed a system start-up, where in a pair of background images are captured, subsequent frames then use the appropriate background image to segment out moving foreground data. Thereafter skin segmentation is done. Here a histogram-based skin classifier assigns each of the RGB pixels in the training set to either a 3D skin histogram or non-skin histogram. The skin segmentation process outputs an image which is ready for detection of colour tapes in the finger to control the mouse cursor movement. The algorithm used here is point pattern matching which is used to find the validity ratio of one database image versus test input image.
- This paper [1] by Nidhibahen Patel and Dr. Selena (Jing) He, gives a complete survey of different techniques for gesture recognition. Some tools which are also involved here for recognition are ANN, HMM and Particle Filtering. Here gestures are being classified as hand and arm gestures, face and head gestures like head rotation, head moving up and down, eye rotation, eyebrows raising and so on. Various technologies are being used such as vision-based (VB) approach, instrumental glove-based (GB) approach and coloured marker approach.

Techniques	Usage	Advantages	Disadvantages
Template Matching	VB and GB	Simplest, Accurate for a small set of postures.	<ul style="list-style-type: none"> • Not for hand gestures. • Does not work for large posture sets.
Feature Extraction	GB	Both postures and gestures, Layered architecture.	Computationally expensive.
Active Shape Models	VB	Real time recognition, Both hand gestures and postures.	Tracks only the open hand.
Causal Analysis	VB	Uses information about how humans interact	Limited Gestures

This paper [3] by Jayshree Katkar and team presents hand gesture recognition-based device control by using microcontroller. Gesture recognition is the interpretation of human motion by computing device. An automatic gesture segmentation algorithm is developed to identify individual gestures in a sequence and the device is controlled by microcontroller. This system uses 2D visual information which is acquired from an ordinary webcam and controls the electrical drives of the device by tracking and recognising the hand gestures. After detecting the hand gesture, the camera will send output to microcontroller via serial communication. Then microcontroller will ON or OFF the device as per the gesture.

3.2 Literature Survey of Speech Recognition

- The paper [11] by Abdulmalik Abdullah Ali Alasadi and Ratnadeep R. Deshmukh presents many of the used speech databases and displays implementation techniques in the feature extraction stage and classification stage for speech recognition systems. It summarises some of the famous and well-known techniques used in various stages of speech recognition system.
- The paper [13] by Hardik Dudhreja and Sanket Shah presents different neural networks used for speech

recognition.

- The paper [12] by Ido Guy examines the logs of a commercial search engine’s mobile interface and compares the spoken queries to the typed-in queries. Here special emphasis is placed on the semantic and syntactic characteristics of the two types of queries. An empirical evaluation is also conducted showing that the language of voice queries is closer to natural language than typed queries.
- In this [10] paper Ashok Kumar and Vikas Mittal proposed a system where a complete perspective on speech recognition describes various processes and summarises various methods used in a typical speech system is represented. Some of the methodologies that are listed are acoustic-phonetic approach, pattern recognition approach, artificial intelligence approach, generative learning approach, discriminative approach and deep learning.
- Arul.V.H and Ramalatha Marimuthu in this [8] paper presents a brief study on speech recognition technology. It further describes the various processing stages implemented and also some primary applications.
- In this [9] paper Pankaj Pathak illustrates what Speech Recognition Technology is, its applications and the future perspective.

4. PROPOSED METHODOLOGY

4.1 Methods & Algorithms for Hand Gestures

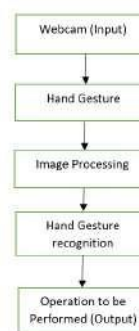


Fig. 1. Architecture of Hand Gesture Recognition

Following methods are used in Hand Gesture Recognition:

1) Image Acquisition:

In this step, an image gets captured from web camera which is connected to the laptop as an input.

2) Pre-processing:

- Convert to grayscale: Image in which the only colours are shades of gray is called as Grayscale image. The reason for differentiating such images from any other colour image is that we need to provide less information for each pixel.
- Removing Noise: A Gaussian function is used to blur an image and the result is called as Gaussian blur. Gaussian blur effect is mostly used in graphics software, typically to reduce the noise of an image.
- Otsu’s Binarization: Otsu’s Method is used to automatically select a threshold value based on the shape of the histogram of the image. This algorithm assumes that the image contains two dominant peaks of pixel intensities in the histogram that is two classes of pixels.

3) Hand Detection:

- Contour: Contour is a curve joining all the continuous points along the borders, having same colour or intensity. Contours are a very useful tools for shape analysis, object detection and recognition.
- Convex Hull: Any region / shape is said to be convex if the line joining any two points (selected from the region) is contained entirely in that region.
- Convexity Defect: Convexity defect can be considered as any

- deviation of the object from the convex hull.
- Extraction of input: After pre-processing on the input image frame mentioned in hand segmentation, hand detection is done. For input purposes, useful information is extracted from the user's hand. For extraction of inputs, contour and convexity defects in the image frame are used.
- Counting Fingers: The count of fingers from the user's hand is extracted in this method. It uses convexity defects for detecting the finger count. The data structure of convexity defect gives the depth of the defects. Many such defects occur in a general hand image because of wrist position & orientation. But there are some defects having far greater depth than others. These are gaps between the two fingers.

4.2 Methods & Algorithms for Speech Recognition

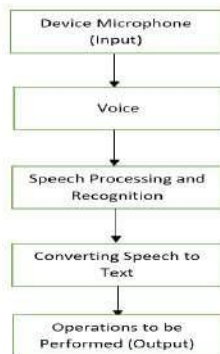


Fig. 2. Speech Recognition Architecture

Speech recognition is the process of converting the input speech into its corresponding transcript. This is very commonly used in the real world. For example, personal voice assistants such as Google's Home Mini, Amazon Alexa, Apple Siri are just some of the popular ones that we know.

Following methods are used in Speech Recognition:

1) Voice Acquisition:

In this step the voice gets captured as an input through the device microphone.

2) Speech Processing and Recognition:

- Speech Processing: The basic aim of speech processing is to provide an interaction between human & machine. Speech processing system has three tasks as follows :

- Speech recognition: Allows the machine to catch the words, phrases and sentences which we speak.
- Natural language processing: Allows the machine to understand what we speak.
- Speech synthesis: Allows the machine to speak.

- Speech Recognition: When a machine understands what humans are speaking it is referred to as Speech recognition. Here for speech recognition we are using Google Speech API in Python.

- For this we need to install following packages:

- PyAudio: With PyAudio, we can easily use Python to play & record audio on a variety of platforms like Apple Mac OS X, GNU/Linux and Microsoft Windows.

We can install it by using `pipwin install pyaudio` command.

- Speech Recognition: For performing speech recognition this library is used. It has support for several engines and APIs which work online as well as offline.

- Speech recognition engine / API support:

- CMU Sphinx (works offline)
- Google Cloud Speech API
- Google Speech Recognition
- Wit.ai
- IBM Speech to Text

- Microsoft Bing Voice Recognition
- Houndify API
- Snowboy Hotword Detection (works offline)

We can install this package by using `pip install speech recognition` command.

3) Google-Speech-API

It accurately converts speech into text using an API powered by Google's AI technologies. We can install it by using the command `pip install google-api-python-client`.

4.3 Hardware and Software Requirements

- Operating System: Windows
- Library: Open CV
- Platform: Anaconda (Spyder IDE)
- Programming Language: Python
- Hardware (Input Device): Webcam for hand gesture and Microphone for speech / voice.

5. IMPLEMENTATION DETAILS

5.1 MODULE 1

In MODULE 1, we would be using Hand Gestures to open and close the Websites.

STEP 1: Enter the website name which you want to open in your web browser. Here we can specify three websites for three different hand gestures.

For example - Google, Yahoo, YouTube, etc.

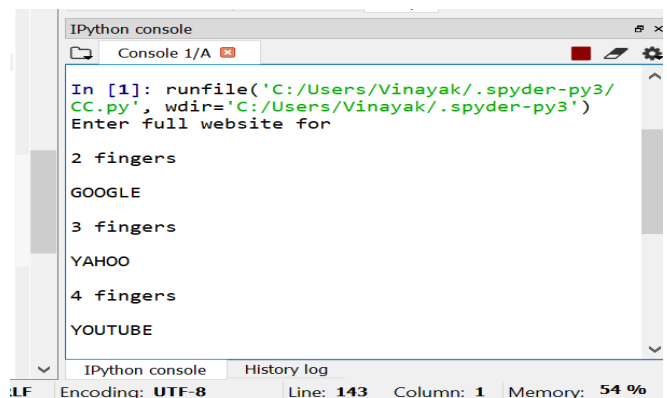


Fig. 3. Website names which we want to open.

STEP 2: Giving the input by showing the hand gestures through the webcam. Here we are specifying a website to a particular hand gesture. So, as we show a hand gesture a website will open in the web browser automatically which is specified to that gesture. We have specified five finger gesture to close the web browser.

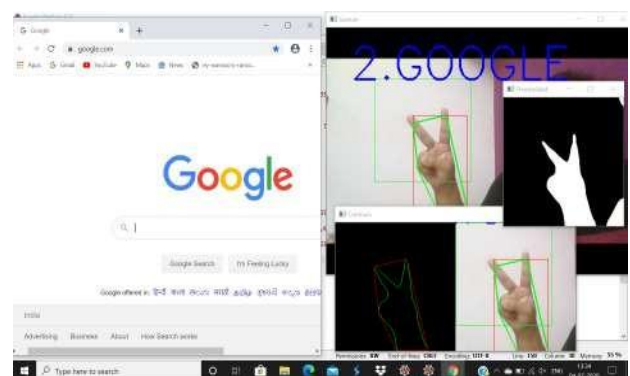


Fig. 4. Hand Gesture recognition to open the 1st tab which is Google.

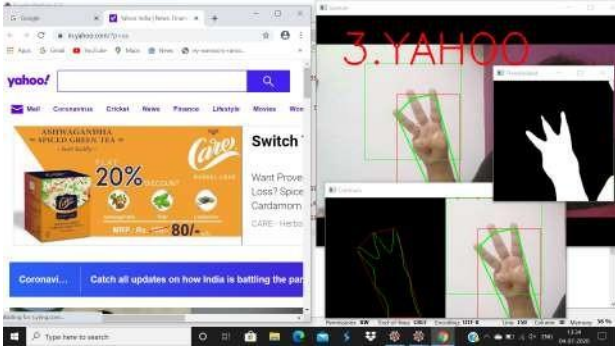


Fig. 5. Hand Gesture recognition to open the 2nd tab which is Yahoo.

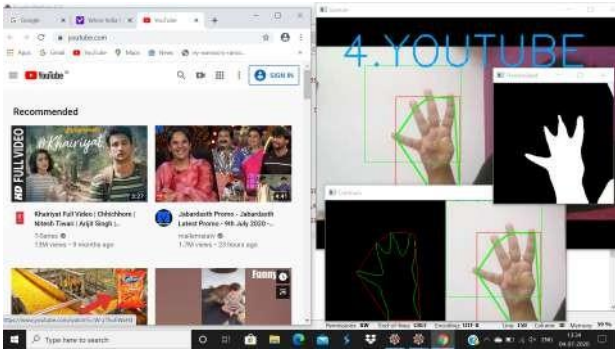


Fig. 6. Hand Gesture recognition to open the 3rd tab which is YouTube.

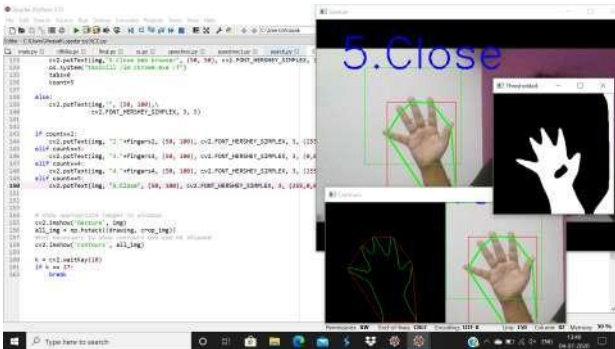


Fig. 7. Hand Gesture recognition to close the Web browser.

which will take speech / voice as input. After that it will search for your request and it will display what you have requested. Here we can say any website name which we want to open. The website will then get opened in the web browser after the searching of request is done. For example: Google, YouTube, Yahoo, Twitter, etc.

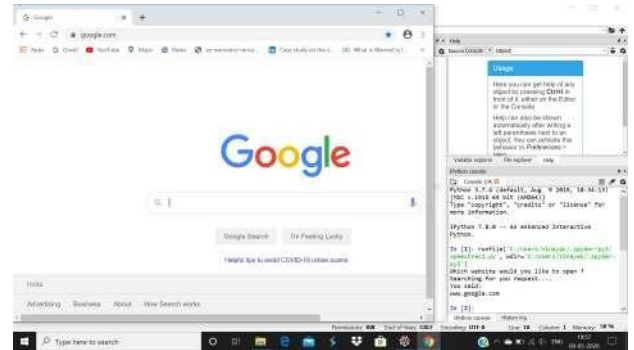


Fig. 9. Result of www.google.com

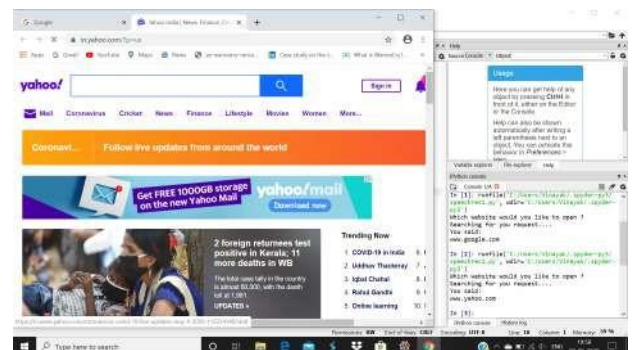


Fig. 10. Result of www.yahoo.com

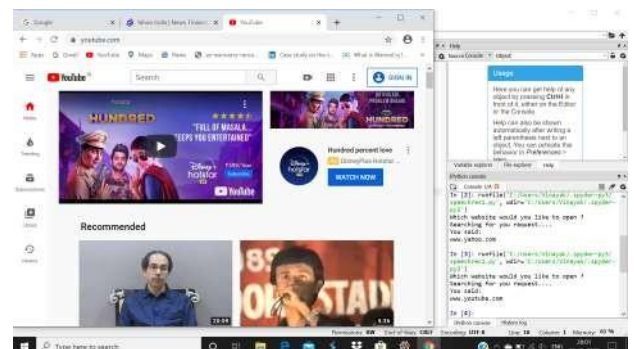


Fig. 11. Result of www.youtube.com

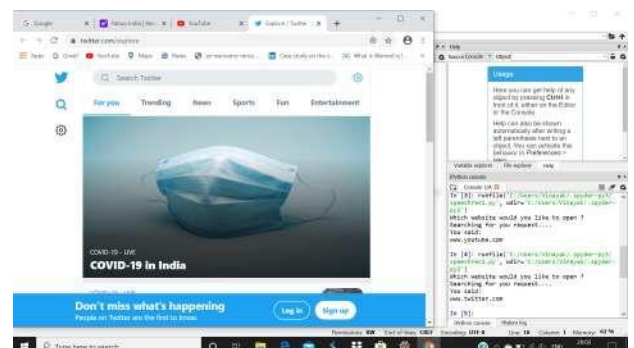


Fig. 12. Result of www.twitter.com

5.3 MODULE 3

In MODULE 3, we would be using Speech to search information on Wikipedia, video on YouTube and music / songs on Gaana.

STEP 1: When we run / execute the program it will ask what do you want to search?

5.2 MODULE 2

In MODULE 2, we would be using Speech to open the Websites.

STEP 1: After running / executing the program it will ask which website would you like to open?

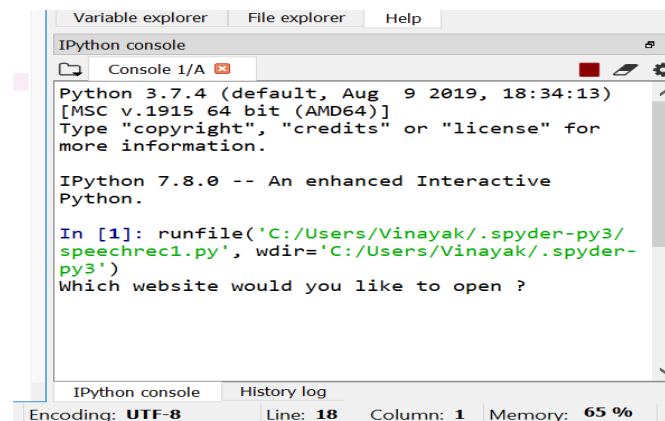


Fig. 8. Which website to open

STEP 2: We need to tell which website we want to open. To enter the website name, we are using the device Microphone

It will provide you three options from which we need to choose:

- Search Web
- Search Video
- Search Music

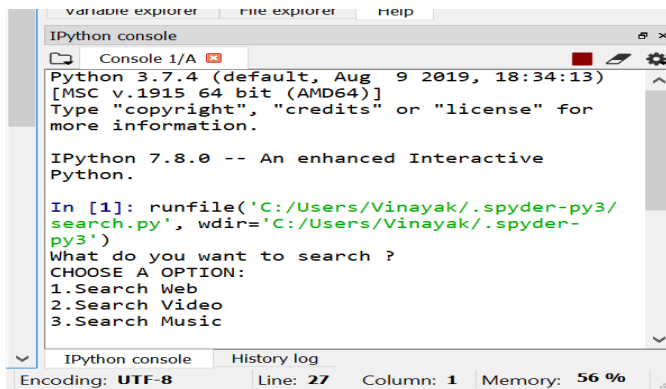


Fig. 13. Search Options

STEP 2: Once we choose any of the options it will further ask you to specify what exactly would you like to search. After we specify our requirement it will search for the same and provide us the output. Following are the results of each of the options:

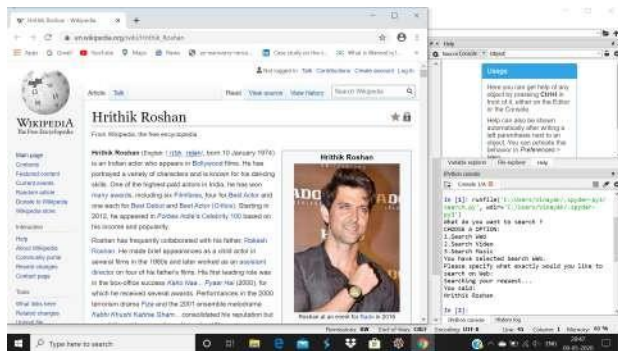


Fig. 14. Result of Search Web

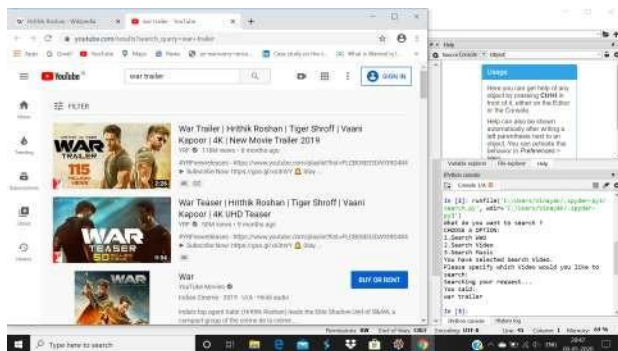


Fig. 15. Result of Search Video

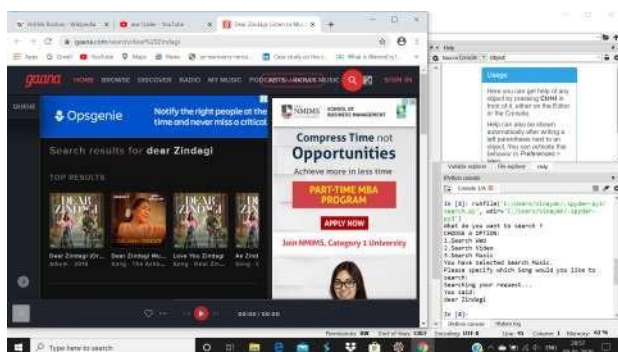


Fig. 16. Result of Search Music

6. APPLICATIONS

6.1 Application of Hand Gestures

- 1) Sign Language Recognition: As we tend to use sign language for interpreting and explaining certain subject during the conversation, it has received special attention. A lot of systems have been proposed to recognise gestures using different types of sign languages. For example, recognized American Sign Language ASL using boundary histogram, MLP neural network and dynamic programming, matching recognized Japanese sign language JSL using Recurrent Neural Network, forty two alphabets and ten words recognised Arabic Sign language ArSL using 2 different types of Neural Network which are Partially and Fully Recurrent neural Network.
- 2) Robot Control: Using gestures to control the robot is one of the most interesting applications. In this field it proposes a system that uses numbering to count the 5 fingers for controlling a robot using hand pose signs. Orders are given to the robot to perform a particular task, where each sign has specific meaning & represents different functions. For example: “five” means “stop”, “one” means “move forward” and so on.
- 3) Graphical Editor Control: This system requires hand gestures to be tracked & located as a pre-processing operation. It uses twelve dynamic gestures for drawing and editing graphic systems. The shapes for drawing are circle, triangle, arc and rectangle. Commands for editing graphic system are copy, move, delete, undo, swap & close, while vertical and horizontal lines are used for drawing.
- 4) Virtual Environments (VEs): Virtual Environments is one of the most popular applications in gesture recognition systems. For communication media systems VEs provide 3D pointing gesture recognition for natural human computer Interaction (HCI) in real-time from binocular views. This system is accurate and independent of environmental changes & user characteristics.
- 5) Recognising Number: Recognising numbers is another important application of hand gesture. We can isolate and recognise a meaningful gesture from hand motion of Arabic numbers from 0 to 9 in a real time automatic system using HMM.

6.2 Application of Speech Recognition

Of-late, speech recognition technology and the use of digital assistants have shifted from our mobile phones to our homes and its application in industries such as marketing, business, healthcare and banking is quickly gaining popularity.

- 1) Workplace: In the workplace speech recognition technology has been incorporated to increase efficiency to perform simple tasks that traditionally needed humans to perform.
- 2) Banking: Speech recognition in banking and financial industry is used for reducing friction with the customer. Voice-activated banking reduces the need for human customer service & lowers employee costs. A personalised banking assistant could result in boosting customer satisfaction and loyalty.
- 3) Marketing: Voice-search can add a new dimension to the way marketers reach to their consumers. Keeping in mind how customers would be interacting with their devices, marketers should look for developing trends in user behaviour and data.
- 4) Healthcare: In the medical profession where seconds are very crucial & sterile operating conditions are a priority, hands-free and immediate access to information can be a boon to the patient’s safety and medical efficiency.

- 5) Internet of Things: Siri's ability to connect to smart lights and smart thermostats makes it seem as though instructing your digital assistant to turn the kettle on is not far off. The Internet of Things is not the futuristic possibility it once was, but rather a relevant development happening around us.

7. CONCLUSION

In conclusion, we can confidently say that hand gestures along with speech recognition would be a great technology to use nowadays for controlling a web browser. Users can use this technology to easily open the websites which they want in the web browser. Thus, a web browser which is controlled via hand gestures and speech recognition is now implemented.

8. FUTURE SCOPE

Since use of hand gestures and speech recognition is increasing in our day-to-day life, we can add many things to it in future. For hand gesture we can implement following things:

- Enhancement of our system to control the whole computer system.
- Development of Gesture-based passwords.
- Implement Gesture based games.

Therefore in future, technologies which will completely remove mouse dependency can be developed with the help of hand gesture- based control system. Speech recognition can be used as an assistant for:

- Switching ON / OFF our light.
- Getting updates about news and weather.
- Reading a book.
- Calling or messaging someone.
- Implementing Voice push notifications to re-engage users with their applications.

Further in future, voice biometrics can also be implemented for authentication as we have evolved from text passwords to numeric PINs to fingerprint and now to facial recognition also.

9. ACKNOWLEDGMENT

First of all, we would sincerely like to thank our college Principal Dr. Sanjay Pawar for giving us an opportunity for doing this project. We are also grateful to our major project coordinator Prof. Sonali Bodekar who provided the required insight and expertise that assisted us to a great extent in our research. We would also like to express our deepest thanks to our guide Prof. Sumedh Pundkar without whose guidance and

suggestions, it would not have been possible for us to complete this project successfully. We were able to enrich our knowledge to a great extent as the process of analysing and doing research on the valuable inputs helped us to explore knowledge, gather useful and valuable information about Hand Gestures as well as Speech recognition and many more things about human computer interaction, giving us a continuous source of inspiration and an unique experience.

10. REFERENCES

- [1]N. Patel and J. He, A survey on hand gesture recognition techniques, methods and tools, 2018.
- [2]V. Shinde, T. Bacchav, J. Pawar, and M. Sanap, Hand gesture recognition system using camera, *International Journal of Engineering Research & Technology (IJERT)*, vol. 3, no. 1, 2014.
- [3]J. Katkar, O. Kahane, V. Jadhav, and P. Jadhav, Hand Gesture Recogni- tion And Device Control, Hand, 2017.
- [4]A. Haria, A. Subramanian, N. Asokkumar, S. Poddar, Shristi and J.S. Nayak, Hand gesture recognition for human computer interaction, *Procedia computer science*, vol.115, pp.367-374, 2017.
- [5]M. N. Devi, M. S. Kuzhali, M. M. Mubeena, Hand Gesture Based Web Navigation.
- [6]R. Z. Khan and N. A. Ibraheem, Hand gesture recognition: a literature review, *International journal of artificial Intelligence & Applications*, vol. 3, no. 4, p.161, 2012.
- [7]S. S. Rautaray and A. Agrawal, Real time gesture recognition system for interaction in dynamic environment, *Procedia Technology*, vol. 4, pp.595-599, 2012.
- [8]A. VH and R. Marimuthu, A Study on Speech Recognition Technology, 2014.
- [9]P. Pathak, Speech recognition technology: Applications & future, *Inter- national Journal of Advanced Research in Computer Science*, vol. 1, no. 4, 2010.
- [10]A. Kumar and V. Mittal, Speech Recognition: A Complete Perspective, 2019.
- [11]R. Deshmukh and A. Alasadi, Automatic Speech Recognition Tech- niques: A Review, 2018.
- [12]I. Guy, Searching by talking: Analysis of voice queries on mobile web search, in *Proceedings of the 39th International ACM SIGIR conference on Research and Development in Information Retrieval*, pp.35-44, 2016.
- [13]M. H. Dudhrejia and M. S. Shah, Speech Recognition using Neural Networks, 2018.