International Journal for Innovative Engineering and Management Research
A Peer Revieved Open Access International Journal
www.ijiemr.org

ASSISTIVE INTERACTION USING GESTURES AND VOICE COMMANDS

Madugula Monalika[1], Rana Ali Khan[2], Shugufta Fatima[3]

[1,2]Student, CSE, Stanley College of Engineering and Technology for Women, Hyderabad, India

[3]Asst. Professor, CSE, Stanley College of Engineering and Technology for Women, Hyderabad, India

Abstract:

To advance is to be human, it's our tendency to explore and move forward and about 50% of the advancements till date have been around the area of communication and accessibility. In the era of keyboards, and keypads, touch screens were a dream, but the human race has gone to those lengths of advancements, such that touch screens are almost a common entity. Now, in the futuristic perspective, voice control and gestures are perceived to be the future! Getting our work done and conveying our messages without even touching. Additionally, while the rest of the world is advancing so much, there ought to be a solution for the disabled community around us, why should they be left behind? With this severe pandemic, we have learnt what goes behind the "touch" and "physical contact", the world has not stopped, no matter what, but temporary solutions where searched and implemented, why not bring out a permanent technology to avoid touch, where it is necessary? Applications of voice and gesture commands are innumerable, from watches to our houses, it has applications everywhere.

Our project will serve to be the first step towards a "no-physical contact" future. We will be using the application of OpenCv library, which is a prime feature present on our desktops, its major role is image recognition, image segmentation and gesture detection analysis. A camera (web-cam) attached to the computer will capture images of the hand, based on the recognized gestures, the text is displayed. This is done using image processing, contour analysis, mathematical algorithms and feature extraction techniques. Similarly for voice control commands we have artificial intelligence-based voice assistants, with an accuracy of 75% which shows us that it has a major scope of development. On a primary note, our project will display the scopes of this improvisation as well. Python is a widely used computer language and perhaps the most accessible, also its increased use in Image Recognition, Detection and Speech Recognition has brought it far off in competition to many other computer languages. As the field of artificial intelligence is constantly developing along with the different Machine Learning algorithms, this futuristic perception seems close to reality. We aim to work with the principles and concepts of Artificial intelligence in our proposed project.

Keywords: voice control, disabled community, OpenCv, Image Recognition, Speech Recognition, Artificial Intelligence.

1 Introduction:

Automation has become the new normal, from homes to cars everything is leaning partially to entirely towards automation. The expected change or let's say advancement in technology over the last few years, have shown up on the insight of the 21st century. Interacting or communicating with our machines is the progressive step towards that advancing goal of a smart world. What do we mean by interacting with a machine? It includes giving the machine some input, and getting an expected output. The input can or cannot be in the conventional way of typing, it can also be voice commands and gestures. Imagine talking to your machine, and getting answers like a personal assistant. Our work can revolutionize the very interaction pattern between us and machines.

An increasingly impactful part of everyday life is how we interact with our technological devices most importantly our computers. Much of the ingenuity stemming from human computer interaction, research focuses only on improving current mainstream devices out there. Only a few modes of Human Computer interaction exist today: namely through keyboards, mouse, touch screens, and other handheld helper devices. Each of these devices has been confronted with their own limitations when adapting to more powerful and versatile hardware in computers.

Over the years, humans have evolved in inventing new technologies for reducing human efforts and saving human life. Physically challenged and elderly people face difficulties while handling objects and hence they need assistance for the same. Thus, if a robotic assistant is developed that can be operated using speech commands would be of immense use. Assistant robots can be used for range of purposes. A robotic assistant that can be controlled using speech commands is developed and it can be used in hospitals, homes, industries and educational institute.

Gestures are employed in our daily life to convey messages, and display emotions. They can also be used to express commands. With rapid advancement in the field of Human Computer Interaction (HCI), it has become possible to gain easy access and control of computer applications using gestures. Using computer vision techniques, it is possible to capture gestures and make interpretations in the form of commands. These commands make it easier for the physically disabled community to control the smart devices, screens, search through the menu.

Digital life assistant-based application which could setup alarm clock, give out the everyday weather information and mails. It automatically sends a mail to some family members or friends added in its list. All the actions said above are followed by voice command or by gestures from the user.

1.1 Objectives

A commonality between the issues of these devices is this: providing an intuitive mode of human computer interaction cheaply without the additive of extra devices. With that said, we felt there were ways we could build an intuitive gesture control mode of human control interaction without the need to necessarily build another device. When tackling this problem of creating a new mode of human computer interaction we knew we could utilize a combination of built-in functions that is typically provided in most computer designs. Specifically, we have exploited the built-in webcam that has become a 6-standard feature in most computers. This feature provides us a way to track and respond user hand free movements and gestures.

No person should struggle with or be limited by technology to complete their day-to-day chores. Yet that isn't always the case for millions who live with some form of disability that limits them. In recent years, a large number of assistive systems for people with special needs have been developed for supporting them in everyday life activities that rely on a variety of different technologies such as speech recognition, gesture recognition, augmented reality (AR), virtual reality (VR) and autonomous robot systems. The aim is to provide an application that includes voice control and gesture recognition that can enable the disabled community to carry out tasks with an ease.

Voice technology can benefit the visually impaired who rely on screen readers, the software that attempts to communicate what is on a display via non-visual means, like text-to-speech, sound icons. The visually impaired who choose to use voice-activated devices, can check their bank balances, turn on lights, play music and much more. Voice assistants are playing a bigger role in the lives of the visually impaired while providing them with new ways to interact with the world.

On the other hand, Gesture recognition provides a whole new level of interaction for the physically disabled. The gesture recognition module understands the command that is being expressed by the user and carries out the required task corresponding to that gesture. One can easily control the mouse operations using

their fingers. Using a combination object detection and recognition, the following project successfully builds a computationally inexpensive static hand gesture recognition system using a simple RGB webcam creating a truly more natural form of human computer interaction.

So finally, the objectives can be listed out as:

- No person should struggle with or be limited by technology to complete their day-to-day chores.
- Voice technology can benefit the visually impaired who rely on screen readers, the software that attempts to communicate what is on a display via non-visual means, like text-to-speech, sound icons.
- Gesture recognition provides a whole new level of interaction for the physically disabled.

1.2 Advantages

Our work revolving around voice assistant and Gesture recognition is very challenging and an interesting task in the field of computer vision, and the technology surrounding both of them is getting mature by the minute, for real world applications. The user gives voice instructions as input and carries out the task in hand. Gesture recognition is also very helpful for the people who cannot reach out to use the mouse and hence can gain a huge benefit out of it. In specific the various advantages are:

- Integration of the voice and gesture control, under a single user interface.
- Business progress especially in the healthcare industry.
- This work in its advanced stages can be a great aid for the visually and audibly impaired.
  It encompasses the depiction of swift migration from 2D to 3D

1.3 Disadvantages

The disadvantages that we noticed with our work is beyond human limitations to fix, which includes the following:

- The user will have to memorize the commands and gestures.
- In case of gestures- discoverability, memorability & fatigue.

2 Existing System:

We have studied 5 different research papers to conclude on the different ideologies that were taken into consideration for the development of the existing systems out there. In general, the most important drawback that we noticed in the existing system is that, there is no system that integrates both of these entirely separate systems, into one system that would encourage them to coexist, while it's essentially necessary Considering the disabled community and the situation post pandemic, for the wide variety of users it's extremely necessary. In general, for specific systems, the drawbacks that we have noticed are as follows:

Speech recognition systems:

- High training time and cost, after that too the uncertainty remains.
- The vocabulary remains limited to the vocabulary of the training set.
- Delayed response is another commonly noted drawback.
- In case of the natural Language Processing, we have noticed it to be reported as inherently ambiguous [1].
- Problem with regards to pauses have been reported with regards to the Discrete Speech Recognition System, which again is a major drawback with the increasing complexities [2].

Gesture Recognition systems:

- To learn, remember and execute the gestures would be a tiring task, leading to fatigue, which has been reported repeatedly for all the existing systems.
- Discoverability of the gestures, keeping in mind the different backgrounds, shapes and similarities between various gestures.
- If we consider the appearance-based approach, we have noticed how it had problems detecting skin like colours and complexes, which is a major background issue [5].
- The fuzzy Hand Posture model solves the background issues with its algorithm to detect the various colors only on the glove, but that again is an Unnatural thing which is practically an extra effort for the user [4].
- Another research work thesis implied that the hand gesture model is susceptible to errors especially with regards to shapes, like squares and circles [3].

3 Proposed System:

The proposed system will provide following features:

1. It always keeps listing for its name and wakes up to response upon calling with the assigned functionality.
2. It keeps learning the sequence of questions asked to it related to its context which it remembers for the future. So, when the same context is mentioned, it starts a conversation with you asking relevant questions.
3. Searching Internet based on user's voice input and giving back the reply through a voice with further interactive questions by machine.
4. Hands-free communication would make the key component of the proposed system as that would ensure its wide-spread usage and application, making it accessible and convenient.
5. Motion history, enables us to keep track on all the gestures and hence makes the gesture control dignified and authentic, as it helps us retrace our steps.
6. The gesture recognition will recognize 4 different trajectories- threshold, black & white, contours and the gesture, this helps to identify the gesture without any difficulty and this also omits most chances of errors.
7. For the execution of this proposed system, no additional devices will be necessary, which would make it even more available and economical. The simplicity of the proposed system compared on par with the complexities of the existing systems is the niche of the proposed system.
8. It gives weather updates, time & date and latest news to the user upon proper delivery of commands by the user.
9. It is also capable of delivering one line jokes upon users command.

4 Architecture

The user can make use of the application in two possible ways. One way is through voice commands and the other way is through hand gestures. When the user gives voice instructions as input, the application processes the voice to remove the background noise and then convert the clear voice into text. Now these text instructions are correspondingly executed.

On the other hand, user can give hand gestures as inputs to control mouse operations. The application processes how many fingers the user is portraying and then capture the hand movement to move the mouse cursor accordingly. The architecture flow in the following figure will help depict our point better.

International Journal for Innovative Engineering and Management Research
A Peer Revieved Open Access International Journal
www.ijiemr.org

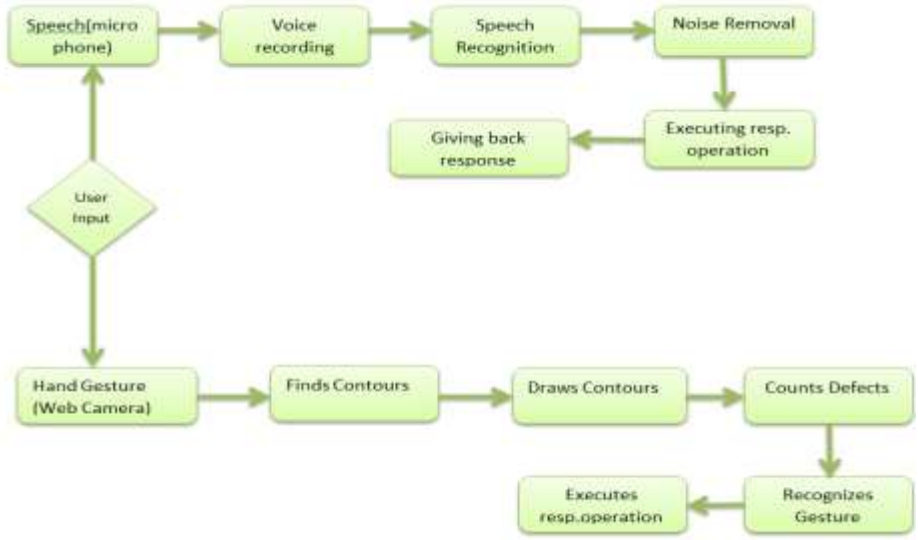

Fig.1. Architecture flow

5 Algorithm

5.1 Algorithm for speech recognition

Step 1: We begin by importing speech recognition library, which is, Speech recognition is the process of converting spoken words to text. Python supports many speech recognition engines and APIs, including Google Speech Engine, Google Cloud Speech API, Microsoft Bing Voice Recognition and IBM Speech to Text

Step 2: We will then initialize the recognizer class. All of the magic in Speech Recognition happens with the Recognizer class. The primary purpose of a Recognizer instance is, of course, to recognize speech. Each instance comes with a variety of settings and functionality for recognizing speech from an audio source.

Step 3: In the recognizer class we are calling the inbuilt speech recognition functions like, recognize_bing(), recognize_google(), recognize_google_clone(). In our project we have included recognise_google() which is used to transcribe our audio files.

Step 4: Audio preprocessing is processing of the audio waveforms to yield a topological map of audio sources as a function of time delay, is done to reform the audio input better.

Step 5: Noise removal which is the process of reducing constant background sounds such as hum, whistle, whine, buzz, and "hiss", such as tape hiss, fan noise or FM/webcast carrier noise. We are using the adjust_for_ ambient_noise(source, duration=1), here the parameters are, source means input, it listens through microphone and duration means it listens for that specified amount of time.

5.2 Algorithm for gesture control

Contour of object is defined by set of points, which describe the edge of object, the outline. For example, the contour of tennis ball is circle. Numerous algorithms for finding contour in digital image were proposed and one of the first was proposed by Theo Pavlidis, where his algorithm is considered as base stone of all others. OpenCV library offers very efficient implementation of contour finding algorithm, which contains additional features like extraction of contours in hierarchy and approximation of found contours. Approximation of line contour with set of points is very handy feature and significantly speeds up further processing of contour. Such

contour, represented as set of points, can be enclosed into n-dimensional polygon, also known as hull. Hull, as the geometrical shape, can be concave or convex polygon. We can say that the hull is convex, when it is not possible to draw a line inside polygon which would intersect its border. If it is possible, then the polygon is not convex and therefore contains convexity defects. These area descriptive properties will significantly help with design of algorithm since human hand does have huge convexity defects between fingers.

Step 1: Creation of Threshold

Creation of threshold image is very crucial for Hand detection. Isolating the foreground from the Background is essential as we want the hand to be the region of Interest.

Fig 2.Thresholding

Step 2: Finding Contours

Contours can be explained simply as a curve joining all the continuous points (along the boundary), having same colour or intensity. The contours are a useful tool for shape analysis and object detection and recognition.

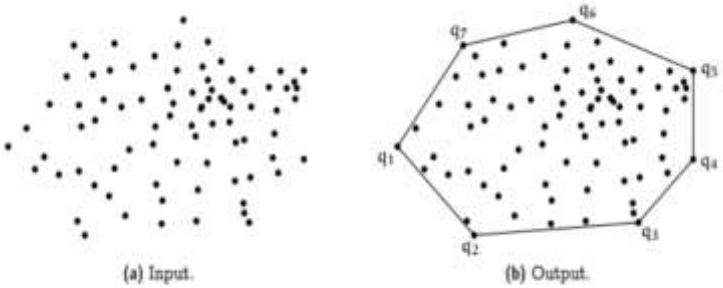

Fig.3.Convex Hull

International Journal for Innovative Engineering and Management Research
A Peer Revieved Open Access International Journal
www.ijiemr.org

Step 3: Creating contours

The Hand is identified using an inbuilt function that finds Contours which OpenCV provides. The function is later then returns an array of co-ordinates of the formation of the Contour.

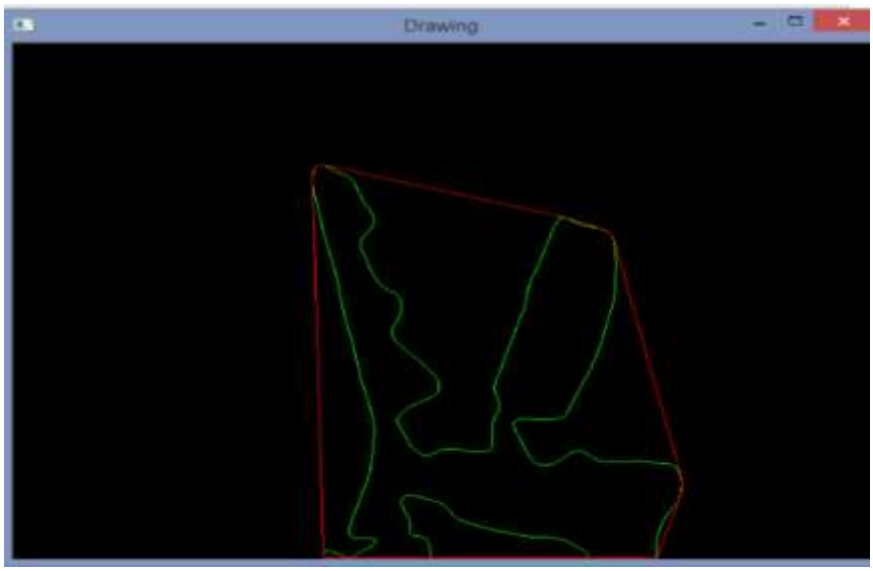

Fig .4.Creating contours

Step 4: Convex Hull and Convexity Defects

The data from the Contour Analysis is later manipulated to obtain an entity known as "Number of Convexity Defects". Convexity Defects are irregularities in the contour. Based on the value of this, we can identify how many fingers are present. This is valuable information as it provides information as to what gesture is corresponding. The number of contour defects is calculated by the following process. We compute a triangle. Let the sides be "a", "b" and "c". This triangle is formed by the starting point of the contour, the ending point of the contour and the farthest point of the contour. (a, b, c respectively) ."a" is computed as follows a = math.sqrt((end[0] - start[0])**2 + (end[1] - start[1])**2) Similarly, b and c are also calculated. Now, using the Cosine rule, If the angle A is less than or equal to 90 degrees, it means that there is a convexity defect. Once there is a convexity defect recognized, a variable named "cnt" increments by one. So, by this algorithm we can efficiently identify how many convexity defects are there.

$$a^2 = b^2 + c^2 - 2bc\cos A \qquad \cos A = \frac{b^2 + c^2 - a^2}{2bc}$$

$$b^2 = a^2 + c^2 - 2ac\cos B \qquad \cos B = \frac{a^2 + c^2 - b^2}{2ac}$$

$$c^2 = a^2 + b^2 - 2ab\cos C \qquad \cos C = \frac{a^2 + b^2 - c}{2ab}$$

Fig.5.Cosine rule for calculating angle A

Step 5: Hand Tracking

The hand-tracking algorithm is simple, efficient and fast so that it can be applied to real-time applications. The algorithm is based on the detection of motion and skin colour. Motion is indicated by the change in the pixel values. The frames are first converted into grayscale images. Then, a frame-differencing algorithm is used to analyse the region where the movement has taken place. The image which has non-zero pixel values in the regions where motion has taken place. A thresholding algorithm gives a binary image with white pixels indicating the region of motion. The threshold of 30 is chosen by monitoring the frame difference in different lighting conditions. The value of 30 gives enough white pixels to track the location of the hand. It is assumed that the only moving object performing the gesture is the hand. Since the video is captured from a regular webcam, random camera noise causes large variations in certain pixel values in successive frames. These variations result in white pixels in the thresholded images.

6 Conclusion

In an era where we are constantly growing and our technology is growing four folds every single second, it's high time we start building something for the aid of those who genuinely need these, for whom these technologies can genuinely make their lives easier. Our work which integrated the voice command and the gesture command, works side by side to help us multi-task and eventually, become self-sufficient, irrespective of the natural factors. To sum up the working of the system, the user can do two things. Firstly, the user can provide voice instructions as input, and gain the required outputs, either via online resources or by offline resources, depending on the requirements. Secondly, the user can provide inputs in the form of gestures, after the gestures are analyzed, they will perform the required mouse navigation. By Integrating both the voice command and the gesture command, we will be taking a small step towards a more advanced and more equipped future. This combined technology will help speed up the building of the smart computer more efficiently and also making it easily available for common users. We live in a world of technical advancements and uncompromised efficiency races! The need to get perfect is what drives the generation around us and the generation to come. Adopting what is new and the need of the hour better be adapted by our business industry. This is super necessary for the advancement of our nation as a whole.

7 Future Scope:

Further research on this work can be done in the following fields to improve user experience and to help with the advancements:

Security Service Field: The traditional security systems are very much dependent on third party human interaction to govern the security systems and that divides the efficiency, hence a more refined way to govern it can be provided with further developments.

Health-Care Field: In case of emergencies where there is absence of human interactions, certain emergency health care devices can be worked upon with our theory, like self CPR kit etc.

Education Field: With the limitations of the traditional classrooms, specially post pandemic, this field has a major area of growth, the very future of the education industry can be revolutionised with future work in our field.

Technological Advancement Field: Smart computers and the VR technology, which have slowly and steadily started taking form of the traditional interfaces, and this technology would make it easily accessible to common people.

Robotic Field: The robotic technology has seen an unforeseen development in the field of accessibility after the development in the Artificial Intelligence and after these further developments in the field of voice and gesture control, we will see an even more amazing development in the same field.

8 References:

1. [1] T. Baudel and M. Beaudouin-Lafon. Charade **:** Remote control of objects using free-hand gestures. Communication of the ACM, 36(7):29- 35, July 1993.
2. [2] T. Baudel and A. Braffort. Reconnaissance de gestes de la main en environnement r~,el. In Actes de "Informatique'93", Interface des mondes re:els et virtuels, pages 207-2t6, Montpellier, 1993.
3. [3] T.F. Coates, C.J. Taylor, D.H. Cooper, and J. Graham. 'IYaining models of shape from sets of examples. In David Hogg and Roger Boyle, editors, British Machine Vision Conference, pages 9---18. Springer-Verlag, 1992.
4. [4] A. Braffort. Reconnaissance et Cornprdhension de gestes, application d la langue des signes. PhD thesis, Universit~ Paris-XI Orsay~ Jmn 1996.
5. [5] B&ard. Vision par ordinateur pour ta r~alitfi augment~e : Application au bureau num(rique. Mas~er's thesis, Universit Joseph Fourier - INP Grenoble, June 1994. \
6. Kishor Kumar Reddy C and Vijaya Babu B, "ISPM: Improved Snow Prediction Model to Nowcast the Presence of Snow/No-Snow", International Review on Computers and Software, 2015.
7. (http://www.praiseworthyprize.org/jsm/index.php?journal=irecos&page=article&op=view&path%5B%5D=17055)
8. Kishor Kumar Reddy C, Rupa C H and Vijaya Babu B, "SLGAS: Supervised Learning using Gain Ratio as Attribute Selection Measure to Nowcast Snow/No-Snow", International Review on Computers and Software, 2015.
9. (http://www.praiseworthyprize.org/jsm/index.php?journal=irecos&page=article&op=view&path%5B%5D=16706)
10. Kishor Kumar Reddy C, Vijaya Babu B, Rupa C H, "SLEAS: Supervised Learning using Entropy as Attribute Selection Measure", International Journal of Engineering and Technology, 2014.
11. (http://www.enggjournals.com/ijet/docs/IJET14-06-05-210.pdf)
12. Kishor Kumar Reddy C, Rupa C H and Vijaya Babu B, "A Pragmatic Methodology to Predict the Presence of Snow/No-Snow using Supervised Learning Methodologies", International Journal of Applied Engineering Research, 2014.
13. (http://www.ripublication.com/Volume/ijaerv9n21.htm)