International Journal of Multidisciplinary Engineering in Current Research
Volume 7, Issue 6, June 2022, http://ijmec.com/

Assistive Interaction Using Gestures and Voice Commands

Mrs Jayashree S Patil1 , Mrs.M.Lalitha2
1 Associate Professor, Computer Science and Engineering Department,
G. Narayanamma Institute of Technology and Science,
Shaikpet, Hyderabad, Telangana, India jshivshetty@gnits.ac.in

2Assistant Professor, Computer Science and Engineering Department,
G. Narayanamma Institute of Technology and Science,
Shaikpet, Hyderabad, Telangana, India mlalitha@gnits.ac.in

Abstract
Now, thanks to advancements in artificial intelligence and speech recognition technology, even those with limited mobility can use computers, laptops, and cellphones. Users are able to do a wide variety of tasks, including as launching apps, sending emails, checking the time and weather, visiting websites, and powering down their devices, all with voice commands. Their conversations are transcribed verbatim and then searched for relevant terms to complete the necessary actions. Paralyzed individuals who are unable to leave their chairs may still use PCs and laptops by pointing and clicking with their hands. The mouse may be controlled by the user's hand motions detected via the use of a variety of hand gestures. Our software uses a camera to record the user's hand movements and gestures, then simulates mouse clicks in response to those movements. This is possible with the use of contour analysis, feature extraction, and hand tracking.

Keywords: Words and phrases like "artificial intelligence," "voice recognition," "hand gesture," "contour analysis," and "hand tracking,".

1. INTRODUCTION

As computing technology evolves, so does the need for standardized methods of communicating between humans and computers. Touch-to-screen technology, which is facilitated by smartphones, is not ideally suited for use in the workplace. The mouse's versatility as a device-control tool is undeniable; nevertheless, it's also very kind to accommodate the really disabled who express a desire in utilizing the mouse for interaction.

In this study, the method used is a camera and voice instructions to record and identify the client's hand movements. This research demonstrates how this method may be used to provide a basic interface for establishing connections. This may be accomplished with the help of Hand Tracking, Contour Analysis, and Feature Extraction.

2. REVIEW OF AVAILABLE WORK

The Worldwide Independent Language Abuse (Storm) initiative is currently being worked on by three DARPA exploratory groups. Information from foreign reports and articles is interpreted and acknowledged by this system. There would be no way to wrap off a conversation on digital personal assistants without mentioning Siri, Google Now, Cortana, and S Voice from Apple, Microsoft, and Samsung, respectively. They allow mobile apps to take voice input from users. The standard Common Language Handling and Discourse Acknowledgement Framework is used in these frameworks. However, if the State Framework

were applied to these programs, response times and user experiences would be enhanced significantly.

Mouse behavior might be manipulated using a variety of hand signals in the past. Many approaches needed sensor-equipped gloves, which increased the already high price and complexity of the application. There was also a scenario when the building could only be built on a solid, noiseless foundation. To properly collect voice instructions, several methods needed a quiet environment. [5]

We want to make the hand-tracking method accessible for real-time use by making it easy to implement and very efficient. The technique allows color analysis and motion detection, which may be used to identify skin tones and discern facial expressions. The photos are first grayscaled, and then a frame-differencing method is employed to pinpoint the active area. Rapid hand motion may be used to scroll the screen, while slow motion in one direction can be used to generate a trajectory. Since the actions are straightforward, less processing time is required. Many scientists have spent the last few decades working to perfect the art of recognizing human gestures. [6]

The Third Pre-Existing System

Section 3.1: Voice Detection

Different voice recognition systems exist to accommodate the wide variety of utterances, speaker models, and vocal capabilities that exist. The difficulties are briefly described below:

Different kinds of speech recognizers:

The many kinds of utterances that may be recognized by a recognizer are used to categorize speech recognition systems. There is a lot of similarity between an isolated word recognizer and a noiseless speech on both ends of the sample window. These word recognizers only support one word input at a time. Similar to single-word expressions, connected words enable for many statements to "run together" with just a brief break between them. As the user types, the computer will figure out what the user is trying to say.

uses normal language while enabling the machine to use natural language. Genuine and unrehearsed expression is the hallmark of spontaneous communication.

Different Speaker Models:

There are two major categories of speaker-based speech recognition systems: speaker-dependent and speaker-independent.

Negatives: Learning Curve, Small Wordlist, Time Lags

Recognition of Gestures 3.2:

Discourse recognition makes use of many crucial methods, including component extraction, acoustic displaying, articulation showing, and decoder. The user sidesteps the program by using techniques designed for a data material device like a receiver. Along these paths, sound waves take the shape of a simple symbol, which the recognizer can decipher before translating into a digital signal. Thereafter, the conversation signal is in the form of electronic pulses. Highlight extraction eliminates several informational foundations, including but not limited to: pitch periodicity, excitation sign abundance, and center repetition. By connecting and progressing data, the decoder does the genuine choice respect acknowledgement of a discourse utterance. [3]

In the twenty-first century, social life and information technology have had an extremely strong link owing to ongoing innovation and advances in computer software and hardware technologies. Consumer electronics product interfaces (e.g., cellphones, games, and infota1inment systems) will grow more sophisticated and feature- rich in the future. Gestures have been used to communicate and

engage with people since ancient times. Humans have the ability to communicate their ideas via body language even before the development of written language. Even in the modern day, many people naturally use gestures, and for the deaf, gestures remain the primary and most natural means of communication. In recent years, gesture control has emerged as a popular new feature on a wide variety of consumer products aimed at humans. People may control these objects more naturally, intuitively, and in the case of the present system, using this method. In this research, we use an adaptive color HSV model and a motion history picture to develop a system for identifying hand gestures in real time (MHI). When utilizing an adaptive skin color model, the effects of lighting, surroundings, and camera may be dramatically reduced, and the reliability of hand motion recognition may be significantly increased.

A computer's keyboard and mouse were among its first input devices, and thus far they have mostly been used for passive data processing. The idea of computer vision arose to give computers the ability to actively acquire data and to broaden the scope of their potential uses. The field of computer vision encompasses a wide variety of subfields:

One Image Is Fed Into The System, and Another Is Produced (improvements filters)

The formula for image analysis is: "Image in, Measurements out" (size, texture, positions, etc.)

• Image Understanding Image in → High-level description out (what is there, what is the link with the surroundings etc) (what is there, what is the relationship with the environment etc.)

The Fourth Proposed System

4.1.1 Voice Activated:

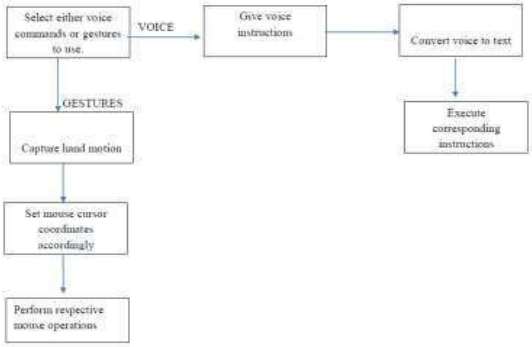
The advantages of the suggested system include: The software is supposed to listen for its name and answer with the defined function when it is invoked. It continues to learn the sequence of questions asked to it in regard to its context, which it retains for future use. As a consequence, anytime the same context is repeated, a discussion develops, with you asking appropriate questions. Using voice instructions to do arithmetic calculations and returning the acquired answer via voice. Using a user's voice to initiate an online search and get a spoken answer using machine learning and natural language processing. By storing information on a remote server, automatic synchronization is always current. Use the Firebase cloud server to make changes to your data in the cloud. Users will be able to connect smart devices to the personal assistant and undertake operations such as turning on and off lights, connecting cellphones, informing the user through push notifications such as email, and so on. Other functions include playing music, setting an alarm, and monitoring the local weather. Setting reminders, checking spelling, and other chores may all be done using only the user's voice.

Recognizing Gestures:

For usage in real-time settings, we developed a simple, powerful, and speedy algorithm for tracking hands. The method relies on the detection of both movement and human skin tone. There was motion shown by the change in pixel values. The first step is to convert each frame into a black and white picture. A frame-differencing technique is then used to the area of interest to determine the nature of the motion. Using a thresholding method based on a set of equations, a binary picture is created with white pixels representing the motion zone. Using observations of frame differences under varying illumination conditions, the value of 30 was arrived at. There are 30 white pixels, which is plenty to follow the hand. These excess white

International Journal of Multidisciplinary Engineering in Current Research
Volume 7, Issue 6, June 2022, http://ijmec.com/

areas may be removed by using the erosion morphological operator twice to the resultant frame. After the region of interest is reduced by an erosion operator, it is enlarged by a series of dilation operations that account for the reduction in size. The attrition process reduces the size of objects by eliminating pixels from their borders in the picture. When the dilation operator is applied to an object, it enlarges it by tacking on pixels to all of its borders. A flowchart depicting the aforementioned procedures may be seen below. To further refine the detection of the area of motion, an adaptive skin detector based on hue thresholds is used. A logic diagram of the adaptive skin detection module. A new set of global thresholds for image filtering may be calculated



using motion detection.
Scrolling when the user's position changes, tracing their path using their global motion vector, and so on are all possible outcomes of this investigation.

One typical use of being able to "read" an image is recognizing gestures. Separated by a transitional period, object detection leads to object identification. The technique of searching for an individual item in a large number of images is called object detection. Object recognition is the effort to identify a trait or characteristic that sets an item apart from its surroundings. To provide an example, object detection can recognize any human hand, but object recognition can tell you what kind of motion it is.

2. SOLUTION

With its ability to record vocal orders and retrieve hand signals, the software offers a means through which the disabled may be assisted. The proposal's user interface has been carefully crafted to be intuitive and simple to use.

3. SYSTEM ARCHITECTURE

The objective of the Voice Recognition component is to build a Google- and email-capable voice assistant. A voice assistant is a kind of digital assistant that delivers a service through an app by recognizing the user's voice and generating synthetic speech. Speech recognition is the foundational technology for voice assistants. They allow computers to understand spoken language. This has the potential to completely alter the landscape for impaired people. Those who are curious in the potential of computers will find this particularly so.

The system may be controlled using gestures made with the help of the gesture recognition module. To communicate information or operate a device, gesture recognition translates the user's actions into digital signals. By using a hand-tracking algorithm, we can roughly pinpoint where the hand is.

Figure-1: The architectural flow of the project

International Journal of Multidisciplinary Engineering in Current Research
Volume 7, Issue 6, June 2022, http://ijmec.com/

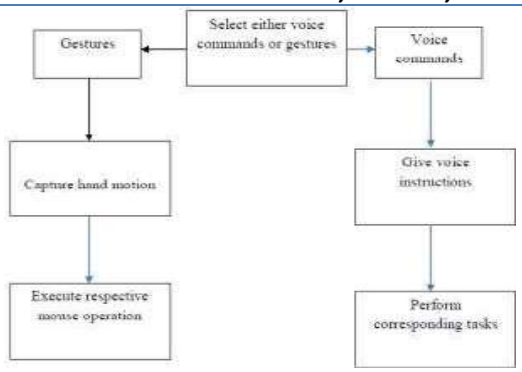


Figure-2: Data flow of the project

The largest cross-section area of any

human hand, regardless of size, is in the middle of the palm, and this is taken advantage of. Using this characteristic and the centroid equation, we may find the mathematical center of the palm. Separating the top from the bottom of the hand is a breeze if you start at the palm and work your way out. Using this strategy, we can generate an accurate algorithm for pinpointing the center of a human hand.

4. RESULTS

4.1.1 Recognizing Your Voice

A user can have a number of things done after the voice module is activated; all he has to do is provide orders using his voice.

The module can recognize speech and filter out unwanted sounds. With this crystal-clear audio, we can now have text written down. Using regular expressions, the algorithm scans the text for terms of contextual relevance. According to the discovered keywords, the corresponding action is taken.

The Recognition of Gestures

In order to recognize gestures, a contour analysis is performed. The contour discovery process is implemented in the OpenCV library in a fairly fast manner, with added features like the extraction of hierarchy-level contours and the approximation of detected contours.

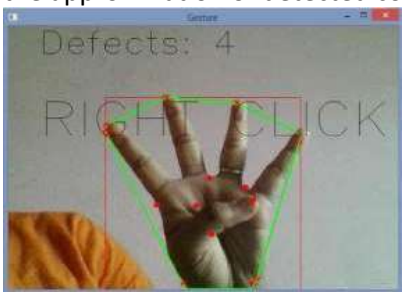


Figure-3: Gesture for Right Click

The ability to quickly and accurately estimate a line's contour using a set of points is a powerful tool. In n-dimensional space, a hull is a shape that represents a collection of points. The hull's polygonal form may be either concave or convex. We say that a hull is convex if and only if it is possible to create a line within the polygon that intersects its boundary using the hull. In this scenario, the polygon lacks convexity and is hence not convex. These area descriptive features will be of considerable help in the construction of algorithms to deal with the huge convexity flaws between the fingers that are characteristic of the human hand.

Figure-4: Gesture for Left Click

Creation of Threshold:

Figure-5: Thresholding

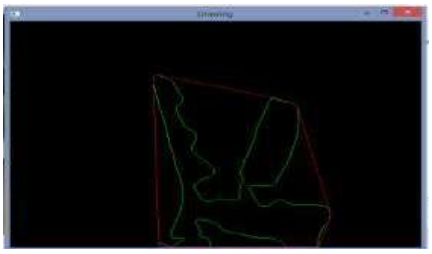
Thresholding It is critical that threshold pictures be created for Hand detection. Since we want

International Journal of Multidisciplinary Engineering in Current Research
Volume 7, Issue 6, June 2022, http://ijmec.com/

the hand to be the RPoI, isolating the foreground from the backdrop is crucial.

How to Use Contours to Determine an Image's Outline:

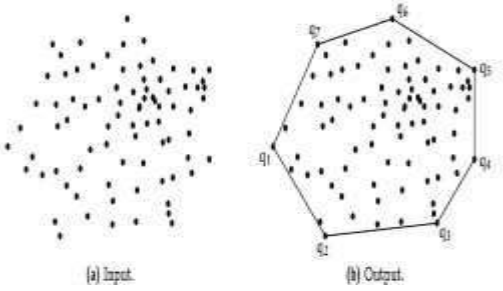
To locate the contours of the hand, OpenCV uses a built-in function called HandRecognize



After the Contour is formed, the method outputs its coordinates as an array.

Figure-6: Finding Contours

Number of Convexity Defects is an entity obtained by manipulating the data from the Contour Analysis to create a convex hull. Disturbances to the contour are known as convexity defects. We can get the total number of fingers from this figure. This is helpful knowledge since it reveals the associated



gesture. Follow this procedure to determine the total number of contour flaws.

Here, a triangle is calculated using the numbers. Assign the letters "a," "b," and "c" to the sides. The three points that make up the triangle are the contour's origin, its terminus, and its furthest extension. (equation (1), (2), (3)). Here's how they're figured out:

The formula for a square root of a number is a = math.sqrt((end[0] - start[0])

Expressed as: **2 + (end[1] - start[1])**2 (1)

B = sqrt(math) (far[0] - start[0])

The formula is: **2 + (far[1] - start[1])**2 (2)

This equation: c = math.sqrt((end[0] - far[0])

The formula is: **2 + (end[1] - far[1])**2 (3)
With the Cosine rule in mind,

cos(A) = (2bc + 2a2)/(2bc + 2a2) (4)

cos(B) = (2a2 + 2c2)/(2a2 + 2b2) (5)

To get cos(C), divide (a2 + b2 c2) by 2ab (6)

In this case, we may compute angle A.

A convexity fault exists if and only if angle A is less than or equal to ninety degrees. If a problem with convexity is found, a counter called "cnt" will increase by one. This approach allows us to quickly count the number of convexity flaws.

Tracking Down the Contour

Basically, a contour is just a line that connects all the spots along a boundary that have the same hue or intensity. It's helpful to have the contours there. tool for shape analysis and object detection and recognition.

Figure-7: Convex Hull

Hand Tracking Algorithm:

Real-time applications may benefit from hand-tracking since it is quick, easy, and accurate. The method relies on the detection of both movement and human skin tone. Motion is indicated by a change in the pixel values. The first step is to convert the frames to black and white. A frame-difference technique is then

used to the area of interest to determine the nature of the motion. Non-zero pixel values may be seen in the picture in the locations where motion has occurred. Thresholding creates a binary picture where the white pixels stand for the area of motion. The number 30 was arrived at by measuring the frame difference under different lighting conditions. There are 30 white pixels, which is plenty to follow the hand.

The hand is taken to be the sole movable part of the gesture. Since the video was recorded using a standard webcam, there may be noticeable changes in the values of certain pixels across subsequent frames due to random camera noise. Due of these deviations, threshold pictures always have some white areas.

4. FUTURE SCOPE

As a result, there is room for development of the work in the following areas to better the user experience and aid the handicapped.

Alternative Computer Interfaces • Medical Applications

• Games for amusement

Technology that automates tasks

Disabled people will have an easier time of it

5. CONCLUSION

The technology for both voice assistants and gesture detection, two challenging but fascinating problems in computer vision, is advancing in real-world applications. In order to accomplish their goals, users submit input in the form of voice instructions. People who are physically unable to use a mouse may benefit greatly from gesture recognition.

REFERENCES

One source is Andreas C. Müller and Sarah Guido's 2018 book for data scientists, Introduction to Machine Learning with Python.

2Cathy Pearl, Designing Voice User Interfaces: Principles of Conversational Experiences, O' Reilly Media, December 2016.

[3] Douglas O'Shaughnessy, Automatic Speech Recognition and Synthesis for Communicating With Computers, IEEE, September 2003.

Learning OpenCV, 1st Edition, O'Reilly Media, September 2008 [4], by Gary Bradski and Adrian Kaehler.

Hand gesture detection and conversion to speech and text, IEEE, 2018. [5] K. Manikandan, Ayush Patidar, Pallav Walia, Aneek Barman Roy.

Advanced Mouse Pointer Control with Trajectory-Based Gesture Recognition. Kabeer Manchanda and Benny Bing. IEEE, 2010.

Reference: [7] Marcus E.Hennecke, K.Venaktesh Prasad, and David G. Stork, Automatic Speech Recognition System utilizing Acoustic and Visual Signals, IEEE, 1996.

Digital Image Processing, Second Edition, by Rafael C. Gonzalez and Richard E. Woods. Prentice-Hall, 2002.

Digital Life Assistant with Automatic Speech Recognition, Seema Rawat, Parv Gupta, and Praveen Kumar, International Conference on Innovative Applications of Computational Intelligence in Power, November 2014.