Chatbot with Gesture based User Input

Rajkumar G[1], Gayathri M[2]

[1]*Student/CSE, SRM Institute of Science and Technology, Chennai-603203*
[2]*Assistant Professor/CSE, SRM institute of Science and Technology, Chennai*

Abstract: The paper describes a approach towards building a chatbot with gesture based input. A gesture table will be described here which is a mapping sequence of the appropriate gestures to the appropriate commands they mean. After successful gesture recognition, the classified or recognized gesture is looked up in the table and is converted accordingly into text and then that text is fed into the chatbot program as input and appropriate command is executed. The paper selects existing known algorithms and aims to explores the overall effects it has and if such a application is suitable for areas such as deaf and dumb sign language or in public galleries, to quickly get required information

1. Introduction

Gesture recognition is a topic in computer science with the goal of interpreting human gestures algorithms. Gestures can be created from any bodily motion but commonly originate from the face or hand. Quite a few approaches have been suggested using cameras and computer vision algorithms to interpret sign language.

Gesture recognition applications in various fields such as automotive, consumer electronics, Transit, gaming, smartphones, defence, sign language translation sectors. Its used to provide a more natural interaction with the UI to save time and make things more intuitive. In entertainment, mainly gaming ,it can be used to add a new way of interaction to attract players. In smartphones, its used for faster unlock. For deaf and dumb, its used to for translating sign language into text.

Gesture recognition has spawned several correlated technology such as touchless interface. Touchless user interface (TUI) is a interface through which a computer can be given commands or operations in form of body motion and gestures without touching a keyboard, mouse, or screen. For example, Microsoft's Kinect is a touchless game interface; This is basically another major application of gesture recognition. Many companies are invested in development of this technology.

Intel corporation is researching how touch multi-factor authentication can help healthcare organizations minimize security risks and at the same time improve clinical efficiency.Here Multi-factor authentication means multiple layers/levels to authorise a particular transaction. Microsoft corporation is researching the use of touchless interaction within surgical settings, allowing images to be viewed and manipulated without any physical contact with traditional computer hardware through gesture recognition.

Gesture recognition is linked to the field of digital image processing. It uses many algorithms and concepts of digital image processing along with concepts and algorithms of AI. Some of these include thresholding, otsu algorithm, anisotropic diffusion, Hidden Markov model, image editing, image restoration etc(belonging to digital image processing), machine learning ,computer vision involving automatic inspection, Assisting humans in identification tasks, Controlling processes, detecting events, interaction, modeling objects, navigation, organizing information.

2. Literature Survey

Rupesh Prajapati, Vedant Pandey, Nupur Jamindar, Neeraj Yadav, Prof. Neelam Phadnis have published a research paper titled “Hand Gesture Recognition and Voice Conversion for Deaf and Dumb”. In it they accept the input in form of Video feed through webcam, they have proposed a method of Create a database of images for training and use PCA for creating a relationship out of its linearly uncorrelated characterstics and then use classfication algorithm like KNN and SVN.[1]

Anchal Sood and Anju Mishra have published a research paper titled “AAWAAZ: A communication system for deaf and dumb-”.They have proposed a sign recognition system based on Harris algorithm for feature extraction and then it is stored in a matrix. This matrix is further used to match the image from the database. The system does have a few limitations. Binary imaging technique doesn’t perfectly seperates discrete objects from backgrounds of similar colors due to inaccurate thresholding algorithm. But the results are efficient.[2]

Ms R. Vinitha and Ms A. Theerthana have published a research paper titled “Design And Development of Hand Gesture Recognition System For Speech Impaired People.” In it they accept input in form of Data glove(consisting of sensors,accelometer and PIC-microcontroller). Gesture recognition is based on hardware approach[3]

Shangeetha, R. K., V. Valliammai, and S. Padmavathi have published a research paper titled “Computer vision based approach for Indian Sign Language character recognition”. In it they accept input in form of Live video feed from webcam. They have also proposed a method of using HIS model and then the feature extraction is done by distance transform method. The results are as follows,The gestures

stored and recognized and coverted to appropriate voice output.[4]

R. Aravind, D.Anbasaran, K.Alice have published a research paper titled "GRS-Gesture Based Recognition System for Indian Sign Language Recognition System for Deaf and Dumb people". In it they accept input in form of webcam. They have proposed a method of gesture recognition using color intensity variation and edge detection and background subtraction. The results were as follows, gesture symbols extracted from the image alone The gaps identified were problems identifying concurrent symbols[5]

Tripathi K. and Nandi have published a research paper titled "Continuous indian sign language gesture recognition and sentence". In they accept input in form of webcam. They have proposed a method using Orientation Histogram and PCA. The results were that highly accurate gesture recognition by reducing dimension features after extraction.[6]

Daniel Simões Lopes have developed a gesture based recognition system as a basis for touchless user interface. In it they accept input in form of Hand gestures and body postures. They have proposed a method of Developed a gesture recognition system where each hand is can make gestures and body postures can also be used. They found out that that using touchless interface is indeed better than relying on 2-D devices as it allows for faster retrieval and greater flexibility in the sensitive surgical enviornment. The gaps that were identified were Precision tasks such as clipping plane and visualization and tagging are better performed with mouse due to not good gesture recognition,[7]

Another interesting method proposed was relying on passive props.3D medical visualization relied on passive interface props. Hardware interface placed in a position where the user needs manipulate the control devices. The result was that the system facilitates a natural interaction between UI and user. The gaps that were identified were that It was not easy to stop or use in several situations, the proposed mechanism is not flexible enough[8]

-Chatbot -Gonzalas Cenelia Article Internet

Gonzalas Cenelia, gives a excellent approach to building a chatbot based on keyword concept. It involves finding keywords in the given user input and match it with a already predefine knowledge base which has keywords and assoiciated responses to it. Revolving around the concept of keyword, things like context of the conversation ,learning from feedback to add knowledge base, keyword ranking etc are also introduced [9]

3. Proposed System:-

Note the high level view of the proposed system in the given diagram.

Input: Imagestream from webcam

Output: Appropriate textual output or execution of a command like search

A. Gesture Recognition Module:-

Input: Imagestream from webcam representing correct gestures

Output:Successfully recognized gesture label(See gesture mapping sequence)

The recognized gestures can be divided into two types.

Counting gestures

-Counting gestures refer to the gestures that indicate numbers like using only your index finger to count 1 ,or any finger for that matter.

Implementation used:Convexity Defect

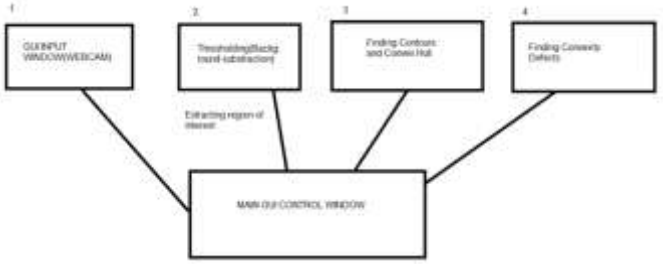
General gestures

-General gestures are the set of gestures that can be classified by a AI model which is trained on a database of images.

Implementation used:CNN model

For each of these types, A separate implementation is provided which is considered more optimum for recognizing those gestures. Using the convexity defect approach, we don't have to train the AI using large number of images for each number representational gesture. Besides the convexity defect approach is very flexible ,it simply counts the number of fingers so we don't have to exact representational gesture like for example using the pinky finger to represent 1 rather than the index finger.(To its recommended to use the normal ones).

I. Convexity Defect Implementation:-



1)Main GUI Control Window

-This window is the main navigation window used to activate all the modules /algorithms for sucessful gesture recognition

2)GUI Input window

-This window activates the webcam and gives you input video feed to manipulate on

3)Thresholding(Background substraction)

As the video feed continues ,every frame is continously getting converted to grayscale image.

A grayscale image is a image without saturation aka color .The only information its pixels have is on the intensity of light falling on that pixel. So basically the image appears in shades of grey.

After this process, thresholding is applied. Thresholding converts the grayscale image to a binary color image which typically are black and white. A value called the threshold point can be manually provided or using an algorithm can be automated . The threshold function will make the pixels whose intensity are greater than the threshold value white and vice versa for black.

In our case,we have use otsu algorithm to calculate threshold point.

4)Finding contours and convex hull and defects

Contours are the areas of the discrete objects that we capture from the image. We assume that the contour which has the maximum area in the binary image will be of our hand, thats why a white screen should be used to hide the background objects.

We calculate then the convex hull which is the smallest polygon that encloses a set of points. Then the convexity defect is defined as the deepest point in the said binary image between two vertices of the convex hull. The actual definition of convexity defect in terms of physical boundaries of the object is given below

Convexity defect is a cavity in an object (blob, contour) segmented out from an image. That means an area that do not belong to the object but located inside of its outer boundary -convex hull.

II. Deep Learning Implementation:

(Contd)

The input consists of 20,000 images . The first Module Image URL loader and reader is a simple algorithm which reads all the appropriate pathnames of the files and converts them into arrays using opencv functions .It extracts all the images required for training and stores the labels for each image in a separate list. One list all the images and the other list contains all the corresponding labels of each image. During the process the images are converted to grayscale and resized to smaller size so that the training phase of CNN model will be shorter. Then its fed into the CNN model. It splits its data into 70% as training and 30% as testing.

(In the below table ,the gestures can be mapped to the appropriate text according to the usage, for example thumb down to what or L to how etc, say if the application is to be used in a deaf/dumb language or a custom gesture system required etc.)

It takes 30 mins approximately for five iterations over the entire database of images for training as well as the validating(testing) to complete. After that model is ready to predict furthur images.

(A sample tableis given if this application is to be used in a museum)

Gesture	Label
Thumb down	How many artifacts are here?
palm(Horizontal)	What is the name of the place
L	Where is it located
Fist(Horizontal)	When was it founded
Fist(Vertical)	What is the most famous artifact here.

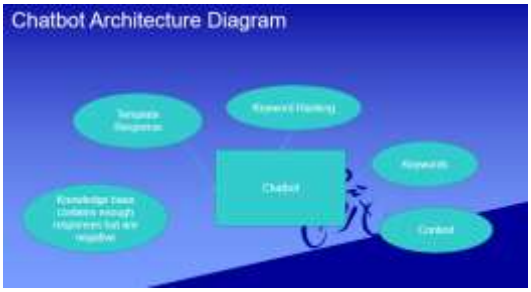
Tools used:

-Python(Anaconda package)

-Spyder(IDE)

-Keras,tensorflow and opencv as python libraries

B.Chatbot



Template Response:- We put up ready-made templates as replies and those templates will have parameters to which will be replaced by the appropriate data relevent to the conversion

Keywords:-The concept on the basis of which the chatbot will be implemented .The chatbot will scan the user input and see if a keyword exists in it. Theses keywords are stored in the knowledge base. And each of the keyword have a set of responses linked to it. So a response is chosen and given as reply

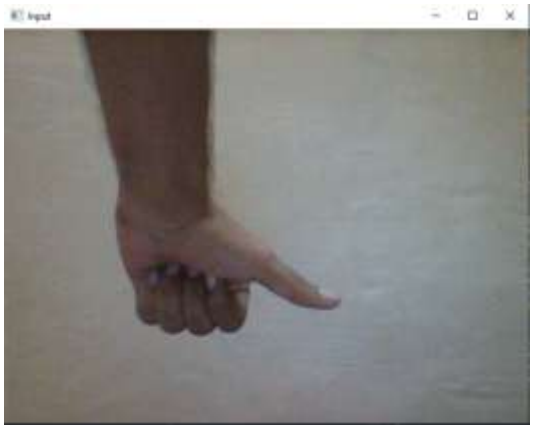
Knowledge base:-The knowledge base is a 2-D array storing a set of keywords in the first column and its associated responses in the next columns.

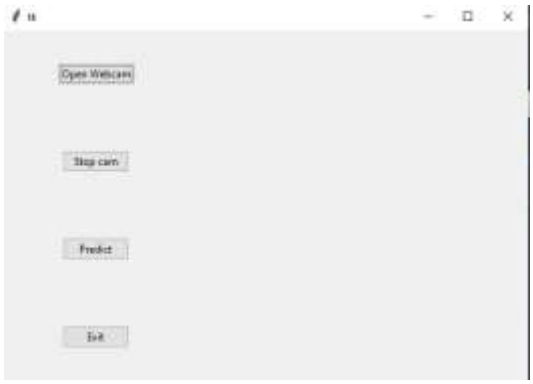
Keyword Ranking:-This is used when there are multiple keyword match and based on the furthur corresponding keyword match and repetitions, the appropriate keyword is selected then the corresponding reply is given

The chatbot will be written in java based module which will accept the input as text which was converted appropriately from the gesture input.

Screenshots:

Deep learning implementation





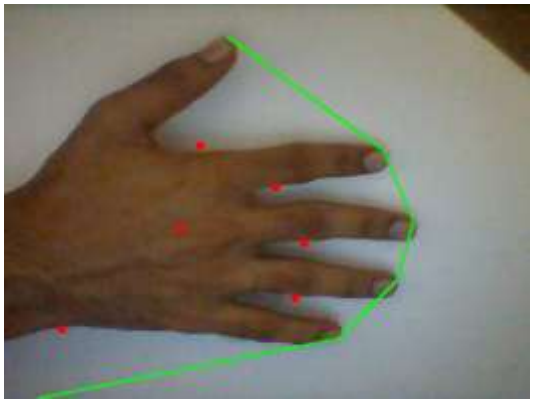
(Note:Predict prints a command line out put.

The image is first captured and then saved

As png file and then is fed into the deep

Learning model)

Convexity defect to count fingers to display numbers



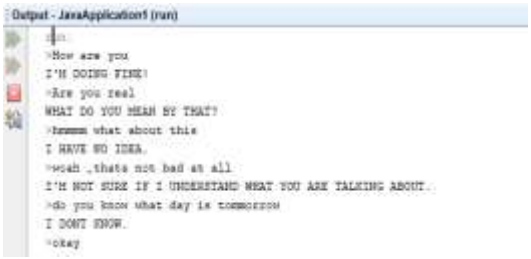
The output is planned to be written to

Text file which will be fed into a java based

chatbot

Chatbot Output:(Separate module)

Example sentences are given as input, it can be customised according to the table created by simple adding those responses to the knowledge base of the chatbot.



References

[1] Rupesh Prajapati, Vedant Pandey, Nupur Jamindar, Neeraj Yadav, Prof. Neelam Phadnis,"Hand Gesture Recognition and Voice Conversion for Deaf and Dumb",IRJET,2018

[2] Anchal Sood and Anju Mishra, "AAWAAZ: communication system for deaf and dumb-".IEEE,2016

[3]Ms R. Vinitha and Ms A. Theerthana,"Design And Development Of Hand Gesture Recognition System For Speech Impaired People.",2016

[4]Shangeetha, R. K., V. Valliammai, and S. Padmavathi,"Computer vision based approach for Indian Sign Language character recognition." Machine Vision and Image Processing (MVIP), IEEE,2012

[5] R.Aravind, D.Anbasaran, K.Alice,"GRS-Gesture Based Recognition System for Indian Sign Language Recognition System for Deaf and Dumb people", International Journal of Trend in Scientific Research and Development - IJTSRD,2018

[6]Tripathi K. and Nandi, "Continuous indian sign language gesture recognition and sentence",Procedia Computer,2015

[7]Daniel Simões Lopes "On the utility of 3D hand cursors to explore medical volume datasets with a touchless interface", Journal of biomedical informatics,2017 Passive real-world interface props for neurosurgical visualization, Hinckley et al.CHI '94,

[9]Otsu Algorithm-(Algorithm used for Thresholding, grayscale to binary),See OpenCv 3.4.2 documentation for algorithm implementation

[10]Peter-Draukr algorithm(For calculation of contours)

[11]Satoshi Suzuki and others. Topological structural analysis of digitized binary images by border following. Computer Vision, Graphics, and Image Processing, 30(1):32–46, 1985.(Algorithm used for finding contours)

[12]CNN(For training the AI model)