International Journal of Research in Engineering and Science (IJRES)
ISSN (Online): 2320-9364, ISSN (Print): 2320-9356
www.ijres.org Volume 9 Issue 7 ǁ 2021 ǁ PP. 61-67

Virtual Talking System without Sensor Image Application for Human Interface Technology

Afshan Taranum[1], Aishwarya M G[2], Anitha[3], Sahil Shariff[4], Prof. Ramya B[*]
[1]*B.E Student Department of ECE VVCE, Mysuru, India*
[2]*B.E Student Department of ECE VVCE, Mysuru, India*
[3]*B.E Student Department of ECE VVCE, Mysuru, India*
[4]*B.E Student Department of ECE VVCE, Mysuru, India*
**Assistant professor Department of ECE VVCE, Mysuru, India*

Abstract:
Abstract—With the impetuous advancement of informatics, human information is unable to bridge the boundaries and human computer interaction is paving the approach for new eras. Here, a real-time human sign gesture recognition using an automated technology called Computer Vision is established. This is a type of noncognitive computer user interface, having the endowment to perceive sign gestures and execute commands based on that. The method is implemented on a Linux framework however can be executed by installing modules for Python on a windows framework too. Creating a desktop application that uses a computer's webcam to capture a person signing gestures for American sign language, and translate it into corresponding text and speech in real time. The translated sign gesture will be acquired in text which is further converted into audio. In this way we are implementing a finger spelling sign language translator. To enable the detection of gestures, we are going to use of a Convolutional neural network (CNN). A CNN is highly efficient in tackling computer vision problems and is capable of finding the desired features with a high degree of accuracy upon sufficient training datasets.

Keywords: CNN (Convolutional neural network), Gesture recognition, Computer Vision, Open CV, KERAS.

Date of Submission: 06-07-2021 Date of acceptance: 19-07-2021

I. INTRODUCTION

The interaction between humans and computers is increasing widely,whereas the domain is witnessing continuous development, with new strategies derived and techniques discovered. sign gesture recognition is one among theforemostadvanced domains in which computer vision and artificial intelligence has helped to improve communication with deaf people but also to support sign gesture-based signaling systems. Subdomains of hand gesture recognition include sign language recognition, recognition of special signal language used in sports, human action recognition, pose and posture detection, physical exercise monitoring, and controlling smart home assisted living applications with hand gesture recognition Over the years, computer scientists have u Different computation algorithms and methods to help solve our problems while easing our lives.

Development of computer and human interaction, and the use of sign gestures in various domains is growing more frequent.

The application of the use of sign gestures can now be seen in games, virtual reality and augmented reality, assisted living, cognitive development assessment, etc. The recent development of sign gesture recognition in different sectors has grabbed the attention of industrytoo for human-robot interaction in manufacturing, and control of autonomous cars.

The main objective of this real-time hand gesture recognition application is to classify and recognize the gestures. sign recognition is a technique in which we use different algorithms and concepts of various techniques, such as image processing and neural networks, to understand the movement of a hand.

In general, there are countless applications of hand gesture sign recognition.For example, for deaf people who cannot hear, they can communicate with their familiar sign language. There are many object detection algorithms been used till now and that helps to detect what the gesture is that each algorithm targets.

II. MOTIVATION

Technology is finding its way towards success in several sectors of human life. Touchless technologies are just around the corner and already penetrates our daily life like smart cars, on-line client support, virtual assistants, smart homes, and virtual reality games.Human communication methods are multiple. Nowadays touchless systems allow us to communicate with a computer by using voice or sign gestures. Voice recognition

and gesture recognition are the types in the computing world and both are evolving at a rapid pace.

III. PROBLEM STATEMENT

Dumb people use hand signs to communicate, hence normal people face problems in recognizing their language by signs madeConsequently there is a requirement for frameworks that perceive the various signs and passes on the data to typical individuals.The normal people can give the voice input and the system replies back with sign output for dumb people.

IV. LITRETURE SURVEY

[1] Ash Jhunjhunwala, Pooja Shah, Pradnya Patil -"Sign Language To Speech Conversion Using Arudino"- A huge population in India alone is of the dumb and deaf people. So the system is working on a hand glove based device which will be used for conversion of sign language (ASL) to speech. The sign language hand glove consist of a simple hand gloves fitted with flex sensors which is being used for the monitoring the amount of bend on the fingers.

[2] Vivek Kumar Verma; Sumit Srivastava; Naveen Kumar-"A Comprehensive Review On Automation Of Indian Sign Language"-Hearing impaired people uses signs to communicate with others.Just like verbally spoken languages, there is no universal language as each nation has its own communicated in language so every nation has their own lingo of communication through signing and in India they utilizes Indian Sign Language (ISL).Over the most recent couple of years, analysts look into the mechanization of ISL.

[3] Vikram Sharma M; Vinay Kumar N; Shruti C Masaguppi; Suma MN; D R Ambika-"Virtual Talking System Image Processing Application For Human Machine Interface Technology"-Every day we see many people who areunable to share their thoughts like deaf, dumb and blind etc. Previously developed techniques are allsensor based and they didn't gave the general solution. This paper clarifies another strategy of virtual talking without sensors and using image processing.

[4] QutaishatMunib MoussaHabeebBayanTakruriHiba AbedAl-Malik -"American Sign Language (ASL)., Recognition Based On HOG Transform And Neural Networks"-The work presented in this paper aims to develop a system for automatic translation of static hand gestures into American sign language. Along with this, in this project they have used Hough transform and neural networks which is trained that is datasets to recognize signs. Framework doesn't depend on utilizing any gloves or visual markings to accomplish the acknowledgment task. and it manages pictures of uncovered hands,which permits the user to interact with the system in a natural way.

[5] The glove based deaf-mute communication interpreter introduced by Anbarasi Rajamohan., Hemavathy R., Dhanalakshmi is a great research. The glove involves of five flex sensors, tactile or material sensors and accelerometer. The system matches the gestures with pre-stored dataset outputs. The evaluation of interpreter was carried out for these letters, _A,, _B,, _C,, _D,, _F,, _I,, _L,,_O,, _M,, _N,, _T,, _S,, _W,,.

[6] As per the Neha V. Tavari A. V. Deorankar Dr. P.N. Authors in this report discuss that many physically impaired and mute, dumb people rely on sign language translators to express their thoughts and to be in touch with rest of the world. This project introduces the image of the hand gestures which is captured using a in-built web camera. Features are used as input to a classification algorithm for recognition. The recognized gesture generate a voice or text. In this system, flex sensor gives unstable analog output and also it requires many electronic circuits and is thus cost effective.

V. OBJECTIVE OF THE PROPOSED WORK

The great challenge lies in developing an economically feasible and hardware-independent system so that physically impaired people can communicate easily as well as normal people can understand it.

- The main objective of our proposed system is to provide speech output using hand gesture signs without using any sensor for dumb people in a smart way.
- We can also give the voice input and the system will give the sign output.

VI. BLOCK DIAGRAM

PROCESSING UNIT

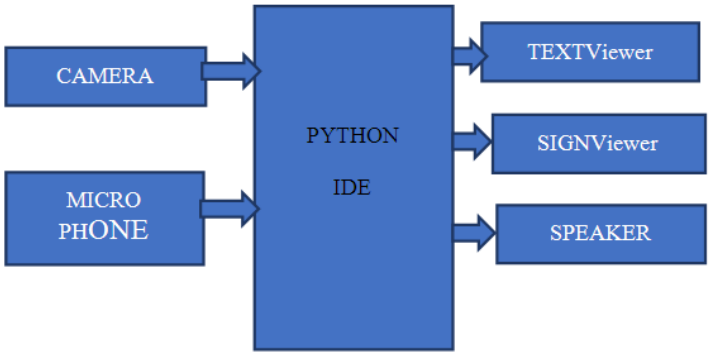

Fig1.block diagram

Input Unit
The input unit of our system consists of Web camera, Microphone, Speaker is used in our design.

Web Camera
Web Camera is the main input device that feeds the processing unit with appropriate quality images. better the quality of an image, and more the accuracy of the converted text, and ultimately the speech.

Microphone
The microphone is a device that translates sound vibration in the air into electronic signals or scribes them to a recording. In this project, the microphone is used as voice input.

Speech Recognition
There are many API for recognizing speech, i.e., converting speech to text. Although there exists CMU Sphinx which works offline, it isn't as accurate as needed. Hence Google speech recognition and Google Cloud Speech APIs have been used here as they are open source and easier to implement in python programs. They can be installed by entering the following command in the terminal *"*pip install Speech Recognition*"* This will install both Google speech recognition as well as Google Cloud Speech APIs. These APIs help us in taking the input when the user wants to retrieve the meaning of a word just by speech.

VII. METODOLOGY

1.SIGN TO VICE CONVERSION
Real-time signal language to textual content and speech translation, specifically: 1. Reading man or woman hand gestures 2. Training the system learning model for image to text or voice content translation 3. Forming words 4. Forming sentences 5. Forming the entire content 6. Obtaining audio output.

A. flow chart of project
The flow chart explains the steps occurring in the system to accomplish

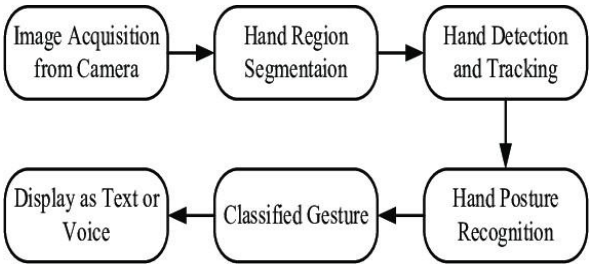

Fig 2 Flow Chart of the Project

The objectives of the project. These steps have been explained in a detail below:
1. Image Acquisition
The gestures are captured through the web camera. The OpenCV video stream is used to capture the entire signing duration. The frames are extracted from the stream and processed as grayscale images with the

dimension of 50*50. This dimension is consistent throughout the project as the entire database set is sized exactly the same.

2. Hand Region Segmentation & Hand Detection and Tracking: The captured images are scanned for hand gestures. This is an element of pre-processing before the image is then fed to the model to obtain the prediction. The segments are containing gestures made more pronounced. This raises the chances of prediction by many folds.

In the fig 3 shown gestures are udes for different signs during training. We have trained only for emergency sign gestures which is helpul during communicate between the deaf and normal people.

Fig.3 Gesture used for different signs during training

3. Hand Posture Recognition
The pre-processed images are then fed to the Keras CNN model. The model that has already been trained, and generates the predicted label. All the hand gesture dataset labels are assigned with a probability. The label with the highest probability is treated to be the predicted label.

4. Display as Text & Speech
The model accumulates the recognized gesture to words. The recognized words are then converted into the corresponding speech using the pyttsx3 library. The text-to-speech result is a simple workaround but is an invaluable feature as it gives a feel of an actual verbal conversation.

B. Convolutional Neural Network

Is for Detection CNN is a class of neural networks that is highly useful in solving computer vision problems. They found inspiration from the actual perception of vision that takes place in the visual cortex of brain. They make use of a filter/kernel to scan through the entire pixel values of the image and make computations by setting appropriate weights to enable the detection of a specific feature. The CNN is furnished with layers like convolution layer, max-pooling layer, flatten layer, dense layer, dropout layer, and a fully connected neural network layer. These layers together make a very warranty tool that can identify features in an image.

C. The CNN Architecture functioning

The CNN model in this project consists of 11 layers. There are 3 convolutional layers. The first convolutional layer,which is based on identifying low-level features like lines, accepts an image with 50*50 size in the grayscale image.16 filters of size 2*2 are used in this layer which results in the form of generation of an activation map of 49*49 for each filter which means the output is nearly equivalent to 49*49*16. A rectifier linear unit (relu) layer is also added to eliminate any negative values on the map and replace them with 0. A max-pooling layer is applied which reduces the activation to 25*25 by only considering maximum values in 2*2 regions of the map. This step raises the probability of detecting the desired feature. This is come after by a second convolutional layer. It is role for identifying features like angles and curves. This layer has 32 filters of size 3*3 which results in the generation of an activation map of 23*23 which means the output is nearly equivalent to 23*23*32.

A max-pooling layer further reduces the activation map to 8*8*32 by finding the maximum values in 3*3 regions of the map. A 3rd convolutional layer is used to find the high-level features like gestures and shapes. The map is straightened to a 1d array of length 64. A dense layer expands the map to an array of 128 elements. dropout layer 64 filters of size 5*5 reduce the input to an output of 4*4*64. A max-pooling layer reseats the map to 1*1*64. Exits irregular guide elements to reduce overfitting. Eventually, a thick layer decreases the guide to a variety of 44 elements which address the quantity of classes.Each class has a comparing likelihood of forecast apportioned to it.The class with the maximum probability is displayed as the predicted sign gesture.

D. Algorithm

Algorithm in Real-time sign language conversion to text and Start
S1: Set the hand gesture histogram to adjust with the skin complexion and the lighting conditions.
S2: Apply data augmentation to the database set to expand it and therefore reduce the overfitting.
S3: Split the database set into train, test, and validation database sets.
S4: Train the Convolutional Neural Network CNN model to fit the database set.
S5: Generate the model report which includes the accuracy, error and confusion matrix so on.
S6: Execute the prediction file – this file predicts the individual sign gestures, cumulates them into words, displays the words as text, relays the voice output. Stop

2.VOICE TO SIGN CONVERSION

It is said that Sign language is the mother language of deaf people. This also contains the combination of sign movements, arms or body, and facial expressions. There are 135 different types of sign languages all over the world. Some of them are American Sign Language (ASL), Indian Sign Language (ISL), British Sign Language (BSL), Australian Sign Language (Auslan), and so on. in this project, we are going to use Indian sign language. This system allows the deaf community to enjoy all sorts of things that normal people do from daily interaction to accessing information.

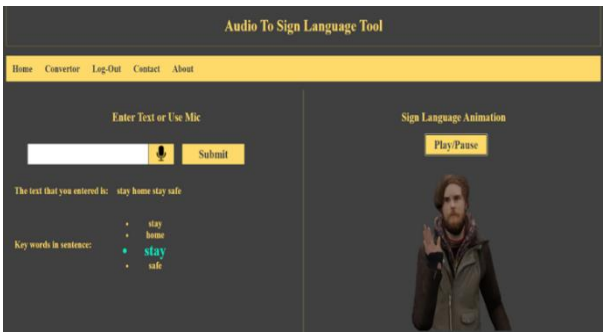

Fig 4 Voice to sign conversion

- This application takes speech as input, then converts it into text, and then displays the Indian Sign Language images.
- The front-end design of the system is designed using EasyGui.
- Speech which is taken as input through the microphone uses the PyAudio package.
- The speech is recognized using Google Speech API
- The text is then pre-processed using with NLP (Natural Language Processing).
- Finally, Dictionary-based machine translation is done.

VIII. IMPLEMENTATION

Output for a given English text is produced by generating its equivalent sign language depiction. The output of the virtual talking system will be a clip of ISL words. The predefined database will be having a video for each and every separate word and the output video will be a merged video of such words.

Fig 5. Speech input

Google Speech-to-Text feature converts audio to text by applying neural network models in an easy-to-use API.

1. Fig 5 showsspeechwhichistakenasinputthroughmicrophoneusesPyAudio package
2. Fig 6 showsthespeechisrecognizedusingGoogleSpeechAPI.
3. Thetextisthenpre-processedusingNLP(NaturalLanguage Processing).
4. Dictionary-based machine translation is done.

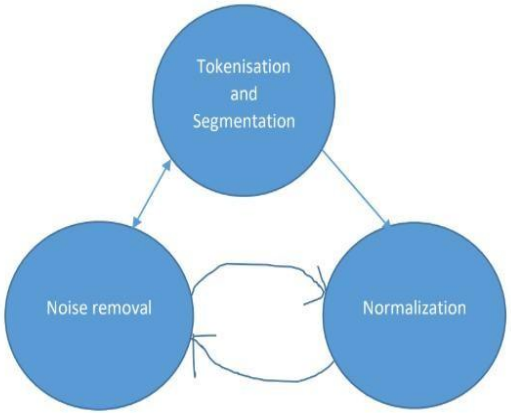

Fig 6 Google Speech-to-text

As we know that Machines can only understand binary language that is 0 and 1 so it is difficult formachinesto understand our language. So, that machine understands human language that is NLP was introduced in this module.NLP that is Natural Language Processing it has the capacity where it measures the content that we have given to the system andunderstands the significance of the voice and text accordingly produces the corresponding output. Text pre-processing comprises of three steps that is Tokenization, Normalization, and the Noise removal as shown in Fig(NLP). Natural Language processing is the mixture of artificial intelligence and computational linguistics. Natural Language processing can do additional functions to our language. We will get our information after producing voice input based on the NLP devices to understand human language. For example like Cortana and Siri.It is a difficult task for the machines to understand our language but with the help of natural language processing, it becomes possible. Here, working operation as shown below as steps:

- Give voice as input to the machine using microphone
- The machine records the audio as input.
- Then machine translates the voice into text and displays .
- The NLP framework parses the text into parts; and understand the context of the conversation and the intention of the person.
- The machine decides which command to be executed, in theview of the aftereffects of NLP.
- Actually, NLP is the process of creating an algorithm that translates text into marking them dependent on the position and capacity of the words in the sentences.

- Human language is changed over genuinely into a mathematical structure. This also makes computers to understand the nuances implicitly encoded into our language.
- When we speak "How Are You" as input into the microphone, the following output pops up as separate letters.

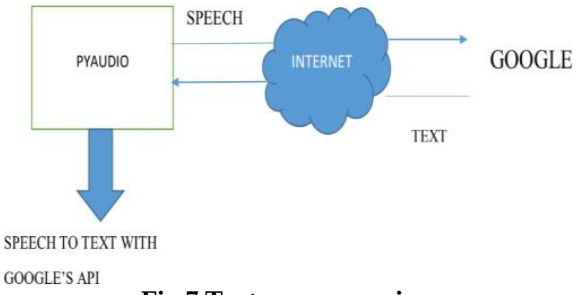

Fig 7 Text pre-processing

IX.ACCURACY OF THE PROJECT

Conversion	Numberof datasets	Accuarcy
Sign to voice conversion	15	85%
Voice to sign conversion	30	95%

X.RESULT

The virtual talking system without sensor image processing application for human interface technology is expected to convert the sign language to voice and text with the voice to sign and proposes an electronic design that can be used for communication between deaf, dumb people with normal people. Web camera is used to take the image of different hand gestures and that will be used as input to the PYTHON. The software will recognize the image and identifies the corresponding text and this text is then converted into a voice using the pyttsx3 library.

XI.CONCLUSION

The sensor-less Virtual Talking systemusing image processing for human interface technology Module is a handy module it provides easy and satisfactory user communication for deaf and dumb people. The module provides two-way communications which help in easy interaction between normal people and with mute people. This module provides an installation for developing a more durable to communicate as sentences.

REFERENCES

[1]. Vikram Sharma M, Virtual Talk for deaf, mute, blindand normal humans, Texas instruments India Educator's conference,2013.
[2]. Concept and Extent of disability in India from Census 2011 and 2001.
[3]. Indian Sign Language Research and Training Centre. A asset for the deaf and dumb people in India
[4]. Q. Munib - American Sign language (ASL) recognition based on Hough Transform and Neural Networks Expert Systems with Applications, 2007.
[5]. P.S. Rajan and Balakrishnan G, Real-time Indian Sign language recognition system to aid deaf dumb people, IEEE 13th International Conference on Communication Technologies, 2011, pp 737-742.
[6]. J. Kim et.al, Bi-channel sensor fusion for an automatic sign language recognition in the 8th IEEE.
[7]. International Conference-2008.