EasyChair Preprint

№ 12802

Hand Gesture Recognition and Voice Conversion for Deaf and Dumb

Eswar Reddy Gudditi, Saikiran Buchiraju,
Dhathri Sriram Pinninti, Sai Lohith Karnati and Kamal Sutaria

March 28, 2024

HAND GESTURE RECOGNITION AND VOICE CONVERSION FOR DEAF AND DUMB

GUDDITI ESWAR REDDY
Computer Science Engineering (CSE),
Parul University, Gujarat, India
gudditieswarreddy@gmail.com

BUCHIRAJU SAIKIRAN
Computer Science Engineering (CSE),
Parul University, Gujarat, India
saikiranjrsk1432@gmail.com

PINNINTI DHATHRI SRIRAM
Computer Science Engineering (CSE),
Parul University, Gujarat, India
srirampinninti918@gmail.com

KARNATI SAI LOHITH
Computer Science Engineering (CSE),
Parul University, Gujarat, India
sailohithkarnati@gmail.com

Dr.KAMAL SUTARIA, Associate Professor
Dept. Of Computer Science Engineering
Parul University, Gujarat, India
kamal.sutaria24554@paruluniversity.ac.in

***Abstract*— Sign language plays an important role in communication between deaf people and normal people. Silent people have difficulty conveying their message to people. Because the average person has not received language education. It is very difficult to convey messages in an emergency situation. Therefore, the solution to this problem is to convert sign language into human voice. There are two main technologies used to detect visual and non-visual gesture or gestures and convert the detected data into speech. Cameras will be used to measure orientation in vision technology. Invisible technique will be used in this project. Most stupid people are deaf. Therefore, people's voices can be transformed into their own signature language.**

I. INTRODUCTION

Nowadays, we always hear that new technologies can improve our lifestyle and make our lives easier. Technology has revolutionized humanity. People are fascinated by technology and have no intention of stopping it. Extensive research is being done on various technologies such as artificial intelligence and smartphones. These studies have led to the emergence of new inventions that make people's lives easier. But there is little research on deaf people. This problem has received less attention than other areas. The real challenge for this special person is the communication between special people and ordinary people. Deaf people always have difficulty communicating with hearing people. This serious problem bothers them and they are hated by people. Deaf people feel that there is no communication. They cannot express their feelings because there is no communication. The HGRVC (Gesture Recognition and Voice Conversion) system can capture and monitor the speech of deaf people to maintain lines of communication with others. Clear recording can be done using a webcam. The image is then converted to different sizes with the help of pre-processing. The aim of this project is to create a system that can convert hand movements into text. The pur-

pose of this task is to place images in a document and convert the images into text by comparing the data. Physical examination Includes hand examination. This method helps bridge the gap between the deaf and the speaker by presenting the text within the text. This article is divided into 5 parts. The first part contains all the research studies. The second is about gesture recognition technology. The third is about how the system works. The fourth is about the consequences of the system, and the last shows the result and reality of our system.

A. *PROBLEM STATEMENT*

It can be used to interact with systems such as navigation, video games, control drones, and medical equipment, among many applications. People with disabilities can interact with the system using available gestures. Classic interactive devices such as keyboards, mice and touch screens can limit the way we use the system. All of these processes require physical contact to interact with the body. Gestures can define the same functionality without the need to physically interact with the interface device. There is a problem in interpreting gestures; The same hand gestures may look different to different people.

B. *SCOPE OF THE PROJECT*

The "Gesture and Speech Conversion for Deaf People" project is comprehensive and has many phases, including collecting movement data and creating programs that can use python libraries to perform speech conversion. Check the description and then we will create software. This is the first step of the development phase of this project. With the help of system diagram, Diagram, then software design and UML diagram, we will make the diagram that will help deaf people to change their new heart with orientation and speech.

C. *OBJECTIVE OF THE PROJECT*

Create software that helps deaf people easily communicate with hearing people and preview historical and real-time data from different sources. Perform engineering to extract relevant features. Explore and compare various machine learning algorithms. Other performance models and choose the best performance model. Create a user-friendly interface to report flood weather and alert authorities and communities

II. MOTIVATION

A. *Background and Related Work*

1. Gesture recognition (HGR) is a computer vision technology that can convert human movements into commands or instructions for human-computer interaction (HCI). 2. It has always attracted attention in many fields for its ability to provide insight. 3.HGR systems generally consist of three stages: hand detection, feature extraction and hand gesture classification.

III. LITERATURE REVIEW

The research paper "A Review of Gesture Recognition in Sign Language Translation" (2013) by A. Mittal, V. Jain and S. Jain provides a comprehensive overview of the handheld tools available for translation. Gesture recognition is a difficult task due to the difference between gestures, the complexity of description, and the need to work on the fly. Advances in information guidance have been made in recent years, thanks to advances in machine learning and deep learning. Advances in information guidance have been made in recent years, thanks to advances in machine learning and deep learning. br¿Deep learning is a good way to recognize gestures in translation. Deep learning algorithms can learn the relationship between hand movements and sign language. The author introduces a deep learning method for hand interpretation that consists of three layers: A convolutional neural network (CNN) layer to remove hand gesture features. Recurrent Neural Network (RNN) layer to model the physical relationship between behaviors. Fully integrated layer to classify gestures based on sign language. This data demonstrates flight performance for signature interpretation using neural networks (CNN). The system was trained on data of more than 10,000 hand gestures taken from the American Sign Language (ASL) fingerspelling corpus. The CNN architecture used in the system has four convolutional layers and two fully connected layers. . Convolutional layers extract motion features, while all layers

split the motions into ASL text. The system is trained using the backpropagation algorithm. The authors evaluated the system with a test of 1000 hand movements taken from the ASL Finger Spelling Corpus. The system achieved an accuracy of 95.21. The authors concluded that the CNN-based recognition function is a good method for interpreting time. They also discuss some limitations of their study and suggest directions for future research. This document describes gesture recognition for the hearing impaired using the Microsoft Kinect sensor. The system was trained using a database of more than 10,000 hand gestures taken from the American Sign Language (ASL) fingerspelling corpus. The system uses various features including hand shape, hand gesture and hand position to recognize gestures. The system also uses a real-time filtering algorithm to remove noise and improve the accuracy of gesture recognition. The authors evaluated the system with a test of 1000 hand movements taken from the ASL Finger Spelling Corpus. The system achieved an accuracy rate of 96.2. The authors concluded that Kinect-based gesture recognition systems are a good method for real-time interpretation for the deaf. They also discuss some limitations of their study and suggest directions for future research. This document describes gesture recognition for deaf people using mobile vision. The system was trained on a database of more than 20,000 hand gestures taken from the American Sign Language (ASL) fingerspelling corpus. The system uses various features including hand shape, hand gesture and hand position to recognize gestures. The system also uses a real-time filtering algorithm to remove noise and improve the accuracy of gesture recognition. The authors evaluated the system with 2,000 hand gesture tests taken from the ASL Finger Spelling Corpus. The system achieved a real-time translation accuracy of 97.2 for the deaf. They also discuss some limitations of their study and suggest directions for future research.

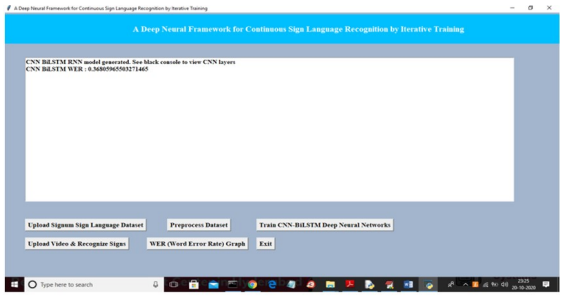
IV. SYSTEM IMPLEMENTATION

A. Training of system

The user needs to enter the number of samples to be stored in the file. The number of samples should be more than 5 for better accuracy. The user must select the folder where the image will be saved. Click ”Start Video” to turn on the webcam to start the database creation process. Click Image Capture to save a series of images to the map. Number of samples. When the number of images equals the number of images captured, a ”Let’s Complete” message will be displayed indicating that file creation is complete.

B. Image Pre-Processing

To improve image quality, image capture enters the first stage. Processing is important to remove objects and background from the image and focus only on that. The original image is then represented in the form of black and white pixels; This means that the image is made binary.



In above screen model is generate and the detected error rate 0.36% and now we can see below black screen to see CNN layers details

Fig. 1. System Architecture

C. Feature Extraction and Recognition

PCA algorithm is used to extract image features. PCA algorithm is applied to the captured images to extract the best images from the data. PCA uses information from the original data, called principal components, to transform the image into a series of distinct lines. Here are the steps to use PCA to extract image features: Step 1: Convert all images into row matrices. Step 2: Measure the average column matrix of the column matrix. Step 3: Calculate the number of changes for each vector group. Step 4: Calculate

the variance matrix. Step 5: Calculate the eigenvalues and mean eigenvalues of the covariance matrix. Step 6: Assets Step 7: Calculate drawings and match project information. After using all the above steps we obtain the reduction of PCA and then calculate the results. After the video is extracted, image recognition analysis is performed using KNN and SVM algorithms. Realize that the gestures are converted into text and then into speech.

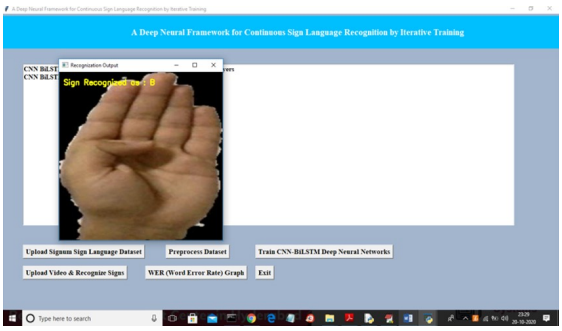
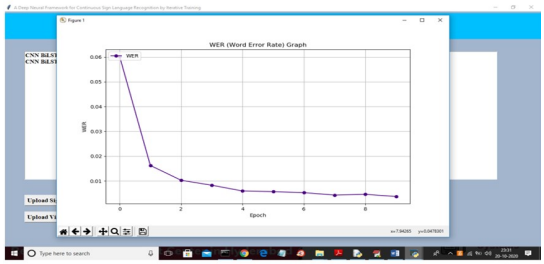


Fig. 2. System Implementation

D. Testing

Software testing is the cornerstone of software security and represents the final review of specifications, design, and coding. Significant increases in the software underlying the system and the costs associated with software failure are incentives to plan through testing. Testing is the process of running a program to find errors. Evaluating software design and other engineering products can be as difficult as the initial design of the product. There are two types of testing: one is black box testing; The other is black box testing. It is a product designed to perform specific tasks, and tests can prove that everything is working fine. Another is white box testing; Understanding the inner workings of products which can be tested to ensure they are produced to specifications and that all by-products are used. The package has been tested using both white box and black box methods. The area and center of each sample circle are checked. Test aims to check every situation and make the right decision. Error handling is done using the special operator. Creating Tests: Tests are activities that can be planned and completed.



In above graph x-axis represents iterative training values and x-axis represents error rate and in above graph we can see when algorithm proceeds further with iterative training then word identifying error rate goes down.

Fig. 3. Result

A software testing strategy should include high-level tests that identify critical functions based on customer needs, as well as low-level tests required to ensure correct use of small numbers. Assessment software is one of the most mature and effective products. Validation is a series of activities that ensure that software is used for a specific task. Verification refers to various activities that ensure that software development can meet the customer's needs. The main purpose of this software is to find errors through testing. To achieve this goal, testing, integration, analysis and testing steps must be planned and executed. Each test step is completed by a test system processes that help generate test data. The level of abstraction determined by the software offered at each testing step . Testing is the only way to ensure software quality and is an umbrella activity, not a separate phase. It is a concurrent work with the software and has its own level of analysis, design, implementation, execution and maintenance.

E. *RESULT*

1. Users can create good information to learn

The only limitation when creating databases is that the background must be white and clear. User should specify model size; default sample size means the number of images in the database.

2. Image capture is done by clicking to capture the image.

3. To get to know the picture, please click on the picture and describe the pictures. The image was processed by KNN and SVM. He did it right.

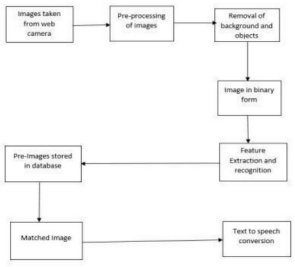


Fig -1: System Architecture

Fig. 4. Flowchart

4. Once the captured image is extracted from the image database, the final result is displayed as a static text. Transformed gestures by feature extraction and image classification.

V. CONCLUSION AND FUTURE WORK

By creating images , he was able to transform gesture recognition and speech for deaf-mutes. This method uses images as input and provides text and speech as output. The system has an accuracy rate of up to 90electronic devices and applications thereby helping the world. Finally, applications that will make the customer's job easier can be created. The main purpose of providing good solutions to customers is to help influence the global community.

REFERENCES

[1] Shinde, Shweta S., Rajesh M. Autee, and Vithhal K. Bhosale. - Real-time two-way communication based on images for deaf and speech-impaired people. – Computational Intelligence and Computing Research (ICCIC), 2016 IEEE International Conference. TSI, 2016.

[2] Shangeetha, R. K., V. Valliammai, S. Padmavathi. - Computer based Indian sign language sign recognition method. - Machine Vision and Imaging (MVIP), 2012 International Conference. IB, 2012.

[3] Sood, Anchal and Anju Mishra. -AAWAAZ: Communication for the Deaf. – Reliability, ICT and Optimization (Trends and Future Directions) (ICRITO), 2016 5th International Conference. IEEE, 2016

[4] Ahire, Prashant G., et al. - Two ways of communication between deaf and hearing people. – Computational Communication Control and Automation (ICCUBEA), 2015 International Conference. IEEE, 2015

[5] Ms. R. Vinitha and A. Sertana. - Design and develop a gesture information system for people with speech disabilities. –

[6] Kumari, Sonal and Suman K. Mitra. - Recognition of people using DFT. – 3rd National Conference on Computer Vision, Pattern Recognition, Image Processing and Graphics (NCVPRIPG), 2011. IEEE, 2011

[7] G. R.S. Murty, R.S. Jayden. (2009). – A review of visual knowledge management, – International Journal of Information Technology and Knowledge Management, Vol. 2(2), p. 405-410.

[8] P. Garg, N. Aggarwal and S. Too fat. (2009). – Vision-based gestures, – Global College of Science, Engineering and Technology, Vol. 49 Ib., 972-977.

[9] Fakhreddine Karray, Milad Alemzadeh, Jamil Abou Saleh, Mo Nours Arab, (2008). Human-computer interaction: a state-of-the-art overview, International Journal of Intelligent Sensing and Intelligent Systems, Vol. 1(1).

[10] Mokhtar M. Hasan, Pramoud K. Misra, (2011). - Luminance matching for gesture recognition using scale normalization -, International Journal of Computer Science and Information Technology (IJCSIT), vol. 3 (2).