Original Article

Voice Based Sign Language Detection For Dumb People Communication Using Machine Learning

Akhilesh Pandey[1], Abhilasha Chauhan[2], Ashish Gupta[3], Vijay karnatak[4]

[1]Assistant Professor, Uttaranchal School of Computing Sciences, Uttaranchal University, India, akhileshmtech10@gmail.com
[2]Assistant Professor, School of Computing, DIT University, Dehradun, India,chauhanabhilasha19@gmail.com
[3]Associate Professor, Computer Science and Engineering, Tula's Institute of Technology, Dehradun, India, ashishgupta.wit@gmail.com
[4]Assistant Professor, School of Computing, University of Engineering & Technology Roorkee, India, vijay.karnatak92@gmail.com
DOI: 10.47750/pnr.2023.14.02.004

Abstract

Around 2.78% of the population in our nation cannot speak (dump). The fundamental way that two people engage is through communication. Since ancient times, people have used speech to interact with one another. Although science and technology have added comfort to human life, there are still some who are less fortunate who are battling to discover a solution that would make communication for them simpler. People who have trouble speaking utilise sign language that is based on hand movements. This programme enables the blind and deaf to interact with seeing people despite their communication challenges. This study suggested a novel method of FFNN (Feed Forward Neural Network) prototype that can automatically recognise sign language to assist normal individuals in more effective communication with those who are hearing, speech, or visually challenged. This system identified of the hand signal feature point extraction given with Feed Forward point extraction given with Feed Forward neural network. Hand Gesture Recognition with Voice Process system using HMM (Hidden Markov model) is used to provide the communication for dump people and the normal people.

Keyword: Gaussian feature extractor, FFNN, Neural Detection, HMM, Video acquisition, Hand Detection.

I. INTRODUTION

Humans interact with one another through sharing their thoughts, ideas, and experiences with others around them. When compared to communication between blind and prehistoric seeing people, communicating with a dumb person presents a significant difficulty. In order to express ideas fluently with a speaker's words, the orientation, movements, facial expressions, body posture, arms, and hands movement. People who are unable to talk use sign languages to communicate with other voice-impaired people as well as with regular those who understand sign language. Interpreters are required to explain the meaning of sign languages to people who can speak but are not familiar with them. It is not always feasible for someone to be present all the time to translate the sign languages, though.

The integration of Artificial Intelligence (AI), Medical science, and computing algorithms are utilized to mimic how people learn, thereby increasing its accuracy. In order to manage image identification effectively, machine learning is essential.Based on the intensity of the pixels in colour or black and white photos, it may determine whether an object is a digital image. Machine learning comes in a variety of forms, including virtual personal assistants, commuter predictions, social media services, video surveillance, and online customer support, among

others. Utilizing supervised learning technologies in this project.A division of machine learning and artificial intelligence is supervised learning. A function or relationship is attempted to be supervised using labelled training data. Based on a collection of input variables, supervised techniques forecast the value of the output variable.

Artificial Intelligence technique are used for procedures which are used to enable computer to show human like intelligent activities such as Speech recognition, Visual perception, Decision-making, and Natural language understanding. Specific application of Artificial Intelligence includes expect system, natural language Processing, speech recognition and machine vision. A system display that can therefore see sign language is shown by the smart interaction on dumb and blind people using hand gesture detection and voice processing, which enables normal people to communicate with groups of people or talk-impaired individuals ever more effectively. This project contains a controller connected to a voice playback circuit, camera, and interface. We may communicate motions by utilising a camera, and for each motion, a speech track is coded. Thus, other conventional people will have no trouble understanding the person who was thwarted. Pre-processing steps used in the recognition of sign language include grayscale conversion, noise reduction, backdrop subtraction, brightness normalisation, and scaling operations. Deaf and mute people's real-time gestures are recorded in a variety of positions.As a database image, this is employed. Real-time gestures are recorded and provided to the pre-processor step as input. The noise component of the colour image is transformed, and the image's undesired background is also removed.Furthermore, techniques enhance for the exact and accurate analysis of the proposed system.

II. LITERATURE SURVEY

Auronno Roy, Md. Tariq Hasan, Md. Mehedi Hasan, and the author create ALAPI: An Automated Voice Chat System in Bangle Language in [1]. This Chabot uses a voice-based system or text-based chat as an input, and the system recognises speech or instructions from the speaker in order to provide the user with an appropriate response or to carry out specified instructions. With the aid of artificial intelligence (AI) and speech recognition technologies, this system can take an audio input and produce a related audio output.

According to Munner et al. [2], they designed a deep learning-based method for recognising hand gestures in sign language. The translation of sign language is a use for hand motion recognition. Multiple deep learning architectures are used in the system's suggested method for dynamic hand gesture recognition, including those for hand segmentation, local and global feature representation, and sequence feature globalisation and recognition. Input pre-processing, feature learning and feature fusion, and classification were the three key stages. The process of the Beidou Navigation System's Global Voice Chat through Short Message Service is designed and tested over BDS SMS in [3], which can transmit voice lasting two seconds in a single 78-byte message. Even if a subscriber is unable to type a single word, they can still use the audio chat option via BDS SMS. This method reduces contact time by more than 90% while enhancing authentication security. The global voice chat successfully augments the outer emergency communication system.

M.A.K. Sudozai and Shahzad Saleem in [4]. In order to make safe chat, audio, and video apps, the author develops a process of profiling from encrypted communications. This process has revolutionised social media. The system has access to a framework for profiling secure voice/video/chat appscan extract the hidden information of different users involved during the process, notifications, devices configuration, status, and other hidden calls without knowledge of the app's communication protocol or its security architecture. Any Android or Windows device can be used for corporate intelligence and criminal investigations.
An MPEG4-based talking head for real-time voice chatting on the Android platform was built in [5] by Jonathan et al... It is a demonstration process. We demonstrate the current level of our work on creating an MPEG-4-based talking head for many platforms in this video. Our most recent achievement is a voice chat software for Android.

Yoko Shinoda, Toru Yamaguchi, Tomova Takatani, Haeveon Lee, Eri Sato-Shimokawara, and Kazuyoshi Wada are included in [6]. The action of For cloud-based chat robots, analysis of category estimation has been conducted while they are being developed and studied. Numerous academics have explored the estimation of categories or topics; yet, voice-based chat systems require a rapid response. A keyword match and prior discourse history are employed.Comparing the analysis of category estimate to handwritten annotation the categorised sentences and

the robot's predicted category are compared. Although the "society" category is hard to describe, "going out" and "music" can be classified as having a 60% ratio.

In [7] An Elicitation Study on Gesture Preferences and Memorability by Abdulmotaleb EI Saddik, Nadia Figueroa, Haiwei Dong, Ali Danesh, and Ali Danesh with the emergence of new depth-sensing technologies for smart televisions, interactive hang-gesture devices have been quickly developing in the direction of a useful HandGesture Vocabulary. Learning multiple vocabularies for various devices can be beneficial. A standardised interactive hand-gesture language is required for users of interactive devices to be able to communicate naturally with one another using hand gestures.The design process for developing such a common hand-gesture vocabulary is discussed. The proposed vocabulary, which resulted from this research, outperforms the current studies in terms of agreement score.

The development of a natural gesture user interface that records and detects hand gestures in real time based on depth data gathered by the Kinect sensor is discussed in [8] by Guillaume Plauffe, Ana-Maria Cretu, and team. The algorithm is suggested to speed up scanning so that the first pixel on the hand contour within this area may be found.The proposed system outperforms most existing solutions for the static recognition of sign digits.According to a comparison of the results with cutting-edge methods, it performs equally for the static and dynamic recognition of well-known signs as well as for the sign language alphabet.

Mart Billinghust, Than Piumsomboon, and Huidong Bai are in [9]. The usage of hand motions with augmented reality interfaces, known as "Hands in Space," is growing in popularity. As more people gain access to this technology, it hacking has been integrated into cheap, publicly accessible gear. In order to facilitate a deeper investigation of gesture-based interaction and interfaces, this work offers researchers and users a straightforward yet efficient technique to implement a variety of onehanded motions.This study presents an easy-to-use yet original method for identifying both static and dynamic one-handed motions. In [10] Hoo Kun and Shi Yuzhuo. Design and implementation of a web-based chat tool is a process that results from the use of cutting tools in a business that specialises in metal cutting, after which the production environment opens up new opportunities for improving the tool management design, methodology, or approach. It was the identification of the present state, exploration condition, current state, exploitation condition, and organisation of the department and other services of cutting tools on lathes and milling machines.There are various drawbacks that have been discovered. Research is especially interesting for CNC workshops and is simple to deploy in businesses with established informatics infrastructure. It is a mechanism used to assess management effectiveness and bring together staff from many departments in a virtual tool management system.

III. Existing System

The addition of Sign Language Recognition (SLR) as an important and promising approach to the current system will help hearing-impaired people communicate more easily. The systems describe the state as the feature representation of a video segment and the action as locating the semantic border. The difference in quantitative performance between the prediction and the true statement.

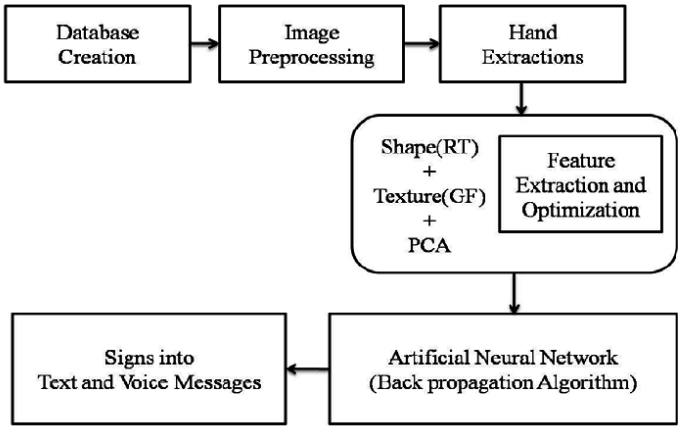

Fig.1: system architecture of the existing system

The identification of video units is essential, but there is a chance of misidentification when a video unit has insufficient details or only partially portrays a sign's gloss. This kind of unit-level misrecognition is the result of a CTC-based strategy used to address incompletion. In addition, there are useless video units inserted between semantic units. Such meaningless units might mistakenly be identified as some glosses while also containing a number of unsettling movements, such as action transitions between nearby glasses and the raising and lowering of hands during videos.Some of the major problems with the existing system are:

- The exact classification of the hand gesture signal is not done.
- There were more process and method used for output.
- The output of this process is a hand signal converted into number or text.
- Accuracy is comparative low at only 80%.

VI. PROPOSED SYSTEM

In the proposal system, a supervised learning approach is enhanced on the hand-based gesture recognition. Here a communicative approach has been lent to the people communication. Feed forward neural network approach provide with the classification of the multiple symbol recognition. The trained features are getting matched with the symbol recognized and the extracted feature points are well known. Matching of the trained feature given with the identification approach of the voice system.

Here, the speech model and the hand-based model are converted using the Hidden Markov Model (HMM) algorithm. A simple chat application is marked by converting the speech signals that match the text. This device offers a display that can recognise sign language, assisting regular people in communicating more effectively with a group of people or in noisy environments. A media filter is used to minimise unwanted noise and remove unnecessary background from images, which can both be added to the image.

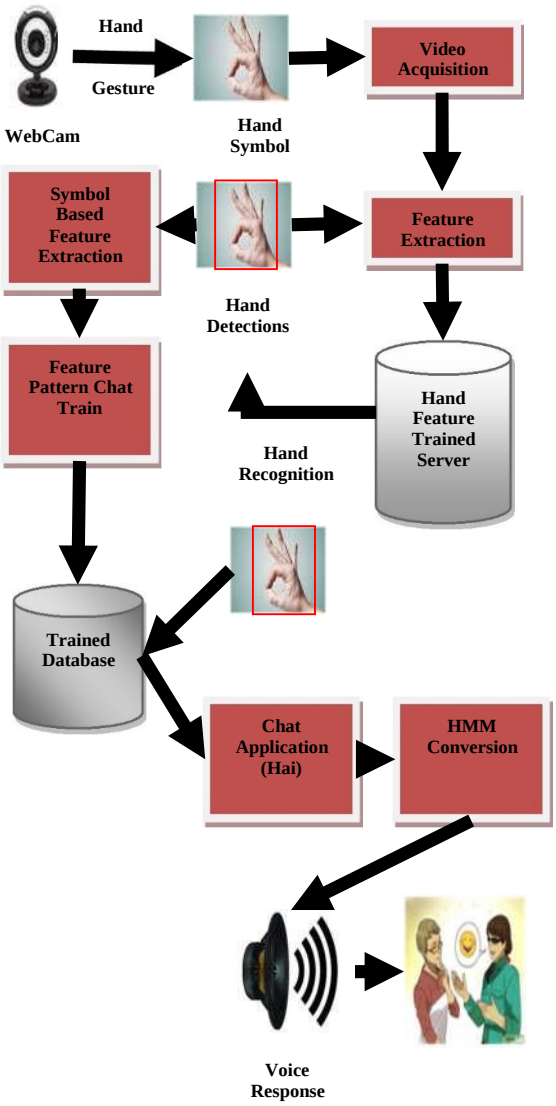

Fig.2: Architecture diagram of proposed Framework

Advantage of the proposed system

- Easy chat application is developed for the dumb and the blind people
- The accuracy of the dump and the blind people
- HMM based voice conversion makes a effective approach of our project

VI. METHODOLOGY

The modules which are used in our proposed system are:

6.1 Video Acquisition:

Video acquisition is the process by which the web cam based accessing system gives out image on light source processing. The light source approach on the process converted as a video and showed out in the screen. These videos will be taken for the further processing or analysis of any other system. Hand gesture recognition can be handled based on station modelling for dynamic gestures by treating them as the results of a stochastic process. The most extensively used visual-based approach is the video acquisition based developed system, which

observes the data obtained by finger and hand movement using a camera. The fundamental drawback of the planned work, however, was its lack of mobility.

6.2 Frame conversion:

Frame conversion is the process where the video will be converted as frames. Frames for an example the video with 60 sec where each will be converted as 60 frames. Each frame will be shown as image with the image the hand symbols are made. Thus, with these image only the extracted feature points are known.

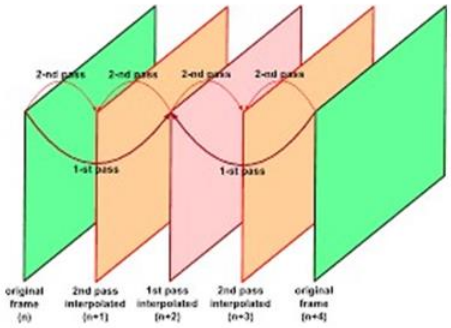

Fig.3: Frame conversation system

The straightforward frame-rate conversion shows the original frames at a different speed than they were originally captured. footage of a time-lapse taken at one frame per minute and replayed at 30 or 25 frames per second The output sampling mode of the output signal parameter, which can either be Frame-based or sampling mode, can also be inherited from the signal, is set by the Frame Conversion block, which also sets the output and sets the output sampling mode to the value of the output signal parameter.

6.3 Gaussian Feature Extractor:

Feature extraction is the pattern of extraction feature points from the images. Feature point is nothing but the outward coordinating points, Morphological points, Texture colour etc. From these feature points the analysis are made exactly for the recognition. With already trained feature dataset will be known to identify the hand gesture from image.

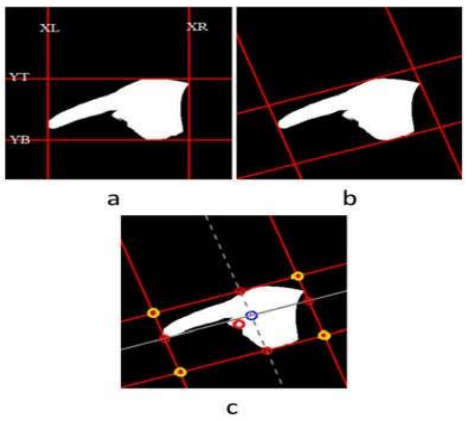

Fig.4: Feature co-ordinator of hand

For the exact extraction Gaussian based feature extractor technique is added. This extracts the correct point of the pixel's points from the image. The given feature formula calculated extracted the peak of the coordinates, probability, total counts with the analyzed processing. The final value executed will go with the training using ML technique.

Gaussian Algorithm

Gaussian(SingularPoints, A)

```
start
        for j=1 to n
 if  akj=0 for all k≥ j, stop; print "A is singular"
            if ajj=0 but akj≠ 0 for  some k>j, switch rows k and j
                for i=j=1 to n
        set zij= aij/ajj
  add _ zij times row j to row I of A
                 next i
                 next j
end
```

6.4 Hand Detection

The ability to recognise hand gestures is essential for human-computer interaction. In this research, we present a novel real-time method for hand gesture identification. To distinguish the hand region from the background, our system employs the background subtraction technique. After that, the finger and plan are segmented to help find and identify the finger. Hand Gesture detection is done with the trained feature matching of hand pattern dataset. The matching feature with the highest probability will give out the exact classification of the different object from the hand. Thus the extracted Gaussian feature technique gives out to match with the hand gesture pattern analysis.

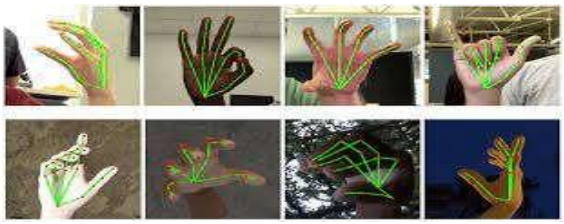

Fig.5: Hand detection system with co-ordinates

An essential component of human computer interaction is vision-based hand gesture recognition technology (HCI). In recent years, keyboard and mouse use have become increasingly important in human-computer interaction.

V. EXPERIMENT RESULT ANALYSIS

In this study, a new technique for hand gesture identification is introduced. The background subtraction method distinguishes the hand region from the surrounding area.

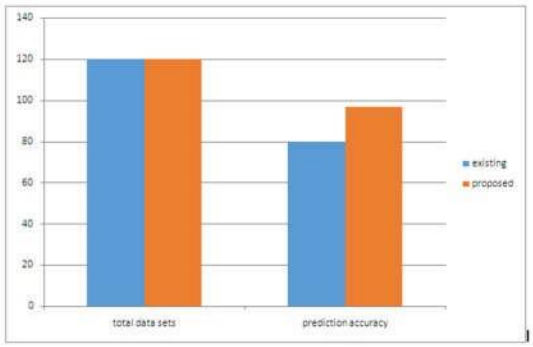

Fig.6:Prediction of Accuracy

Segmented skin covers the palm and fingers. The fingers in the hand image are found and identified on the basis of segmentation. The simple rule classifier is used to recognise hand gestures. A straightforward rule-based classifier recognises hand motions. 1300 photos of hands are used in the evaluation of our method's effectiveness. The results of the experiments demonstrate that our method works well and is appropriate for real-time applications. On a set of hand gesture images, the suggested method performs better than the state-of-the-art Gaussian feature extraction.

Table 1 the accuracy graph measurement of existing and proposed solution

Sign Language	Existing system Accuracy(%)	Proposed system Accuracy (%)
He	88	98
She	87	97
Okay	83	97
A	78	94
B	76	96
1	84	94
7	87	97
4	85	95

The outcome of hand detection has a significant impact on how well the proposed strategy performs. The performance of hand gesture identification will suffer if there are moving items whose hue is comparable to that of the skin. These things will appear in the results of hand detection. However, the hand may be distinguished from the background using machine learning methods. ToF cameras offer the depth data necessary to enhance hand detection efficiency. The robustness of hand detection may therefore be increased in further efforts by addressing the challenging background problem using machine learning techniques and ToF cameras.

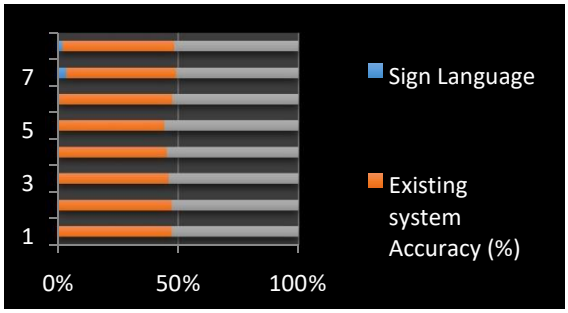

Fig.7: prediction of features for existing and proposed system

The trained features are getting matched with the symbol recognized and the extracted feature points are well known. Matching of the trained feature based on the voice system.Here Hidden Markov Model (HMM) algorithm proposed to convert the voice model and the hand-based model.

VII. CONCLUSION AND FUTURE ENHANCEMENT

This project's primary goal is to aid the deaf and the homeless. The hand gesture recognition and voice conversion system can assist those who have difficulty speaking quickly in real-world interactions. The project is a helpful tool for speech challenged and partially paralysed individuals that fills the communication gap between people. Communication plays a key role in expressing what one wants, in an emergency, or to communicate one's sentiments. One of the helpful methods for facilitating communication between the deaf and mute populations and mainstream culture is sign language. Although sign language can be used to communicate, the recipient must understand sign language, which is not always achievable. Hence our project lowers such barriers. In the future enhancement the detection given accuracy of level which can be made and change with good feasible accuracy system.

VIII. REFERENCE

[1] Md. M. Hasan, A. Roy, and Md. T. Hasan, "ALAPI: An Automated Voice Chat System in Bangla Language," 2021 International Conference on Electronics, Communications, and Information Technology (ICECIT). IEEE, Sep. 14, 2021. doi: 10.1109/icecit54077.2021.9641323.

[2] M. Al-Hammadi et al., "Deep Learning-Based Approach for Sign Language Gesture Recognition With Efficient Hand Gesture Representation," IEEE Access, vol. 8. Institute of Electrical and Electronics Engineers (IEEE), pp. 192527–192542, 2020. doi: 10.1109/access.2020.3032140.

[3] Y. Wei, H. Liu, J. Ma, Y. Zhao, H. Lu, and G. He, "Global voice chat over short message service of Beidou navigation system," 2019 14th IEEE Conference on Industrial Electronics and Applications (ICIEA). IEEE, Jun. 2019. doi: 10.1109/iciea.2019.8834348.

[4] M. A. K. Sudozai and S. Saleem, "Profiling of secure chat and calling apps from encrypted traffic," 2018 15th International Bhurban Conference on Applied Sciences and Technology (IBCAST). IEEE, Jan. 2018. doi: 10.1109/ibcast.2018.8312271.

[5] J. Ozsvath, B. Eros, A. Toth, and R. Kunkli, "An MPEG-4 based talking head for real time voice chatting on Android platform," 2017 8th IEEE International Conference on Cognitive Infocommunications (CogInfoCom). IEEE, Sep. 2017. doi: 10.1109/coginfocom.2017.8268272.

[6] E. Sato-Shimokawara, Y. Shinoda, T. Takatani, H. Lee, K. Wada, and T. Yamaguchi, "Analysis of category estimation for cloud-based chat robot," 2016 25th IEEE International Symposium on Robot and Human Interactive Communication (RO-MAN). IEEE, Aug. 2016. doi: 10.1109/roman.2016.7745147.

[7] H. Dong, A. Danesh, N. Figueroa, and A. El Saddik, "An Elicitation Study on Gesture Preferences and Memorability Toward a Practical Hand-Gesture Vocabulary for Smart Televisions," IEEE Access, vol. 3. Institute of Electrical and Electronics Engineers (IEEE), pp. 543–555, 2015. doi: 10.1109/access.2015.2432679.

[8] G. Plouffe and A.-M. Cretu, "Static and Dynamic Hand Gesture Recognition in Depth Data Using Dynamic Time Warping," IEEE Transactions on Instrumentation and Measurement, vol. 65, no. 2. Institute of Electrical and Electronics Engineers (IEEE), pp. 305–316, Feb. 2016. doi: 10.1109/tim.2015.2498560.

[9] M. N. A. Nor'a and A. W. Ismail, "Integrating Virtual Reality and Augmented Reality in a Collaborative User Interface," International Journal of Innovative Computing, vol. 9, no. 2. Penerbit UTM Press, Nov. 28, 2019. doi: 10.11113/ijic.v9n2.242.

[10] S. Yuzhuo and H. Kun, "Design and realization of chatting tool based on web," 2013 3rd International Conference on Consumer Electronics, Communications and Networks. IEEE, Nov. 2013. doi: 10.1109/cecnet.2013.6703312.