SIGN LANGUAGE TO SPEECH TRANSLATION USING MACHINE LEARNING

Mrs.Aerpula Swetha, Assistant Professor, Department of CSE, swethayadav501@gmail.com

Vamja Pooja, B.Tech, Department of CSE, poojavamja11@gmail.com

Vundi Vedavyas, B.Tech, Department of CSE, vundivedavyas@gmail.com

Challa Datha Venkata Naga Sai Kiran, B.Tech, Department of CSE, dvnsaikiran@gmail.com

Sadu Sravani, B.Tech, Department of CSE, sravanisadu1999@gmail.com

ABSTRACT: Sign language is an incredible advancement that has grown over the years. Unfortunately, there are some drawbacks that have come along with this language. Not everyone knows how to interpret a sign language when having a conversation with a deaf and mute person. One finds it hard to communicate without an interpreter. To solve this, there is a need for a product that is versatile and robust. There is a need to convert the sign language so that it is understood by common people. So here the aim is to get the deaf and mute people more involved to communicate and the idea of a camera-based sign language recognition system that would be in use for converting sign language gestures to text and then to speech. There are major techniques available to detect hand motion or gesture and then converting the detected information into voice such as CNN algorithm.

1. INTRODUCTION

Sign Language has become a most common method of communicating to those people who cannot speak. It is a language that uses the hand motions to express alphabets and words. People who are using the sign language were recorded just in china alone . It exceedsupto 80 million in total and especially those people will always have a problem of communicating with each others who can’t understand the sign language. Vision method has become the popular method used for sign recognition in the past decades. It is a system which uses a camera to sense the information that has been obtained through finger motions. It is the most commonlyused visual-based method. It has been a tremendous effort and has been gone into the development of vision-based sign recognition systems through worldwide. Vision-based gesture recognition systems can be divided into direct and indirect methods. In earlier days for recognizing hand motion, vision based technique is used. But in this method the environmental effect in the recognized image is high and another disadvantage is they have to show their hands to in front of the camera.Here flex sensor is used for detecting the hand motion and convert it into voice.

This topic has got less attention as compared to other sectors. The Main challenges that this special person facing is the communication gap between -special person and normal person. Deaf and Mute people always find difficulties to communicate with normal person. This huge challenge makes them uncomfortable and they feel discriminated in society. Because of miscommunication, Deaf and Mute people feel not to communicate and hence they never able to express their feelings. HGRVC (Hand Gesture Recognition and Voice Conversion) system localizes and tracks the hand gestures of the mute and deaf people in order to maintain a communication channel with the other people. The detection of hand gestures can be done using web camera. The pictures are then converted into standard size with the help of pre-processing. The aim of this project is to develop a system that can convert the hand gestures into text. The focus of this project is to place the pictures in the dataset and with database matching the image is converted into text. The detection involves observation of hand movement. The method gives output in text format that helps to reduce the communication gap between deaf-mute and people.

Fig.1: Example figure

2. LITERATURE SURVEY

Two way communicator between deafandmutevs common people:

One of the most precious gift of nature to human beings is the ability to express himself by responding to the events occurring in his surroundings.Every normal human being sees, listens and then reacts to the situations by speaking himself out. But there are some unfortunate ones who are deprived of this valuable gift. This creates a gap between the normal human beings and the deprived ones. This application will help for both of them to communicate with each other. The system is mainly consists of two modules, first module is drawing out Indian Sign Language(ISL) gestures from real-time video and mapping it with human-understandable speech. Accordingly, second module will take natural language as input and map it with equivalent Indian Sign Language animated gestures. Processing from video to speech will include frame formation from videos, finding region of interest (ROI) and mapping of images with language knowledge base using Correlation based approach then relevant audio generation using Google Text-to-Speech(TTS) API. The other way round, natural language is mapped with equivalent Indian Sign Language gestures by conversion of speech to text using Google Speech-to-Text(STT) API, further mapping the text to relevant animated gestures from the database.

Orientation sensing for gesture-based interaction with smart artifacts:

Orientation sensing is considered an important means to implement embedded technology enhanced artifacts (often referred to as ‘smart artifacts’), exhibiting embodied means of interaction based on their position, orientation, and the respective dynamics. Considering artifacts subject to manual (or ‘by-hand’) manipulation by the user, we identify hand worn, hand carried and (hand) graspable real world objects as exhibiting different artifact orientation dynamics, justifying an analysis along these three categories. We refer to orientation dynamics as ‘gestures’ in an abstract sense, and present a general framework for orientation sensor based gesture recognition. The framework specification is independent of sensor technology and classification methods, and elaborates an application-independent set of gestures. It enables multi sensor interoperability and it accommodates a variable number of sensors. A core component of the framework is a gesture library that contains gestures from three categories: hand gestures, gestures of artifact held permanently and gestures of artifact that are detached from the hand and are manipulated occasionally. An inertial orientation sensing based gesture detection and recognition system is developed and composed into a gesture-based interaction development framework. The use of this framework is demonstrated with the development of tangible remote controls for a media player, both in hardware and in software.

Automated Speech Recognition Approach To Continuous Symbols Generation

The work described in this paper is with an aim of developing a system to aid deaf-dumb people which translates the voice into sign language. This system translates speech signal to American Sign Language. Words that correspond to signs from the American sign language dictionary calls a prerecorded American sign language (ASL) showing the sign that is played on the monitor of a portable computer. If the word does not have a corresponding sign in the sign language dictionary, it is finger spelled. This is done in real life by deaf for words that do not have specific signs like for proper names. Hidden Markov Model (HMM) is used for recognition of speech signal from the user and translated to cue symbols for vocally disabled people. The proposed task is a complementary work to the ongoing research work for recognizing the finger movement of a vocally disabled person, to speech signal called “Boltay Haath”. The proposed AISR system integrated with Boltay Haath system could eliminate the communication gap between the common man and vocally disabled people and extend in both ways.

Finger motion detectionfor sign language recognition

Computer recognition of sign language is an important research problem for enabling communication with hearing impaired people. This paper introduces an efficient and fast algorithm for identification of the number of fingers opened in a gesture representing an alphabet of the American Sign Language. Finger Detection is accomplished based on the concept of Boundary Tracing and Finger Tip Detection. The system does not require the hand to be perfectly aligned to the camera or use any special markers or input gloves on the hand. Index Terms—Boundary Tracing, computer access for disabled, finger detection, image processing, sign language recognition.

Recognition of arm gestures using multiple orientation sensors: gesture classification

We present a gesture recognition algorithm from Euler angles acquired using multiple orientation sensors. This algorithm is a part of a system for controlling unmanned aerial vehicles (UAVs) in the presence of manned aircrafts on an aircraft deck. After exploring multiple approaches to arm gesture recognition, we investigate a real-time arm gesture recognition system using the IS-300 Pro Precision Motion Tracker by InterSense. Our work consists of (1) analyzing several gesture recognition approaches leading to a selection of an active sensor, (2) gesture modeling using Euler angles, (3) low-level gesture characterization, and (4) model-based gesture classification algorithms. We have implemented and tested the proposed real-time arm gesture recognition system in a laboratory environment with a robot that represents an UAV surrogate.

3. PROPOSED SYSTEM

An interpreter is required to translate the meanings of sign language to people who are able to communicate verbally. However, it is not always possible for someone to be present all the time to interpret sign languages and not everyone is capable of learning them.Also there are few methods to convert the sign language which often use Kinect as the basic system to get inputs and work on them for conversion.Kinect methods are complicated in so many aspects due to which processing would be complicated.

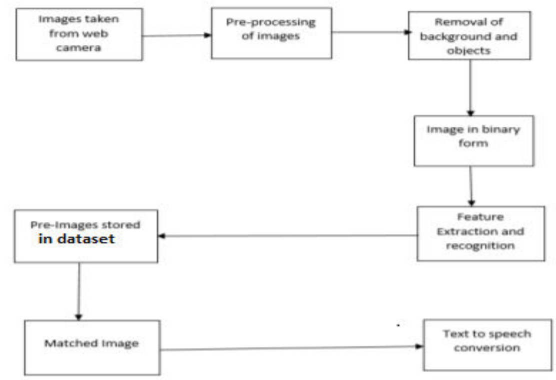
DISADVANTAGES:

- A lot of parameters must be fixed and modified to get accurate results.

The main objective is to design a solution that is intuitive and simple which simplifies the communication for the majority of people with deaf and mute people. Here first we try to recognize the hand signs or gestures and display the correspondent word. The first phase involves capturing the gesture using a webcam along with pose estimation library. The webcam captures the image and image is processed in tensor-flow utility to get converted into text and then to speech.

ADVANTAGES:

- The accuracy rate of translation is high and is understood by common people which will help them to communicate without any interpreter.



.Fig.2: System architecture

4. IMPLEMENTATION

- We are using deep learning Convolution Neural Network to train hand gesture photos and then this trained model can be used to predict those trained hand gesture from webcam because the author of the proposed study used the SVM technique but Python SVM is not accurate in identifying hand gesture.

CNN ALGORITHM:

A neural network type called a convolutional neural network, or CNN or ConvNet, is particularly adept at processing input with a grid-like architecture, like an image. A binary representation of visual data is a digital image.

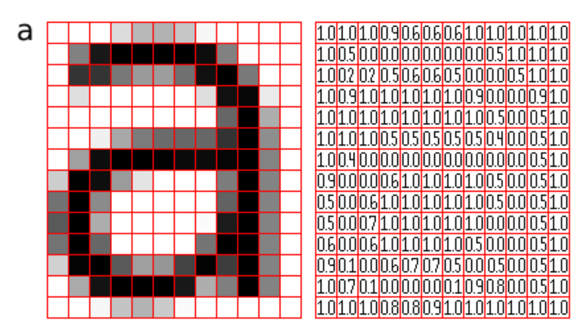


Fig.3: Representation of image as a grid of pixels

The moment we perceive an image, the human brain begins processing a massive amount of data. Every neuron has a distinct receptive field and is coupled to other neurons so that they collectively cover the whole visual field. Each neuron in a CNN processes data only in its receptive field, similar to how each neuron in the biological vision system responds to stimuli only in the constrained area of the visual field known as the receptive field. Lines, curves, and other simpler patterns are detected initially by the layers, followed by more intricate patterns like faces and objects. One can enable sight to computers by employing a CNN.

Convolutional Neural Network Architecture:

Convolutional, pooling, and fully connected layers make up a CNN's standard three layers.

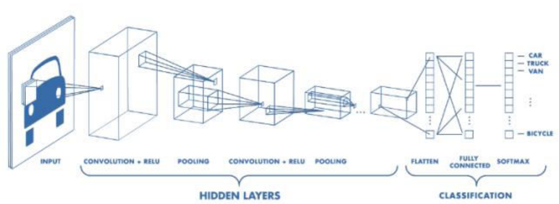


Fig.4: CNN architecture

Convolution Layer: The CNN's fundamental building piece is the convolution layer. It carries the majority of the computational load on the network. This layer creates a dot product between two matrices, one of which is the kernel—a collection of learnable parameters—and the other of which is the constrained area of the receptive field.

Pooling Layer: By obtaining a summary statistic from the surrounding outputs, the pooling layer substitutes for the network's output at specific locations. This aids in shrinking the representation's spatial size, which lowers the amount of computation and weights needed. Each slice of the representation is subjected to the pooling operation separately.

Fully Connected Layer: Similar to standard FCNN, all neurons in this layer are fully connected to all neurons in the layer above and below it. Because of this, it can be calculated using a matrix multiplication followed by a bias effect, as per usual. The representation between the input and the output is mapped with the aid of the FC layer.

5. EXPERIMENTAL RESULTS

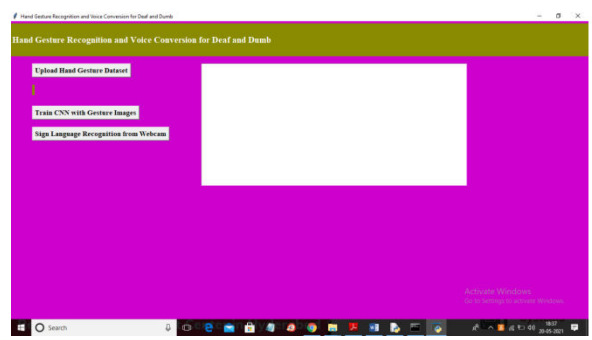


Fig.5: Home screen

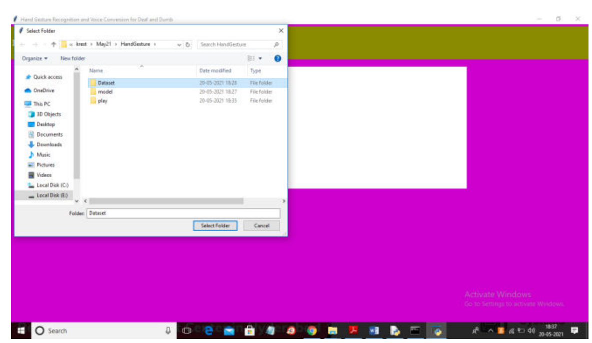


Fig.6: Upload hand gesture dataset

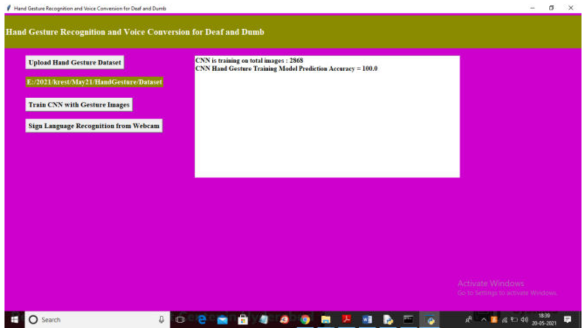


Fig.7: Train CNN gesture images

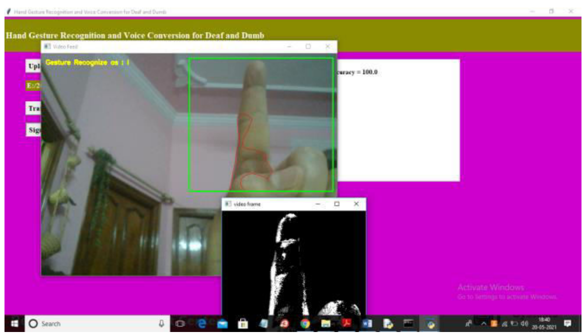


Fig.8: Sign language recognition from webcam

Here, you simply need to demonstrate the motion as it is shown on the screen above. If you adjust your hands, you may receive an incorrect prediction; but, if you fix your gesture, the prediction will be correct. The modules listed below are executed for each prediction when the project is run.

- Webcam picture extraction, conversion to binary or grayscale, and background removal
- Extract image features
- Recognize and play audio

6. CONCLUSION

Hand Gesture recognition and voice conversion for deaf and mute people was successfully executed using image processing.The method takes hand gesture image as input using a webcam and gives text and speech as an output.Implementation of this system gives up to 90% accuracy and works successfully in most of the test cases.In future it would even be possible to grow our dataset and get more gestures translated to speech. Thus using this system, even deaf and mute people will be able to communicate at ease.

REFERENCES

[1] Shinde, Shweta S., Rajesh M. Autee, and Vitthal K. Bhosale. "Real time two way communication approach for hearing impaired and dumb person based on image processing." Computational Intelligence and Computing Research (ICCIC), 2016 IEEE International Conference on. IEEE, 2016.

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

[3] Sood, Anchal, and Anju Mishra. "AAWAAZ: A communication system for deaf and dumb." Reliability, Infocom Technologies and Optimization (Trends and Future Directions)(ICRITO), 2016 5th International Conference on. IEEE, 2016.

[4] Ahire, Prashant G., et al. "Two Way Communicator between Deaf and Dumb People and Normal People." Computing Communication Control and Automation (ICCUBEA), 2015 International Conference on. IEEE, 2015.

[5] Ms R. Vinitha and Ms A. Theerthana. "Design And Development Of Hand GestureRecognition System For Speech Impaired People."

[6] Kumari, Sonal, and Suman K. Mitra. "Human action recognition using DFT." Computer Vision, Pattern Recognition, Image Processing and Graphics

(NCVPRIPG), 2011 Third National Conference on. IEEE, 2011.

[7] S. F. Ahmed, S. Muhammad, B. Ali, S. Saqib, and M. Qureshi, "Electronic Speaking Glove for Speechless Patients A Tongue to," no. November, pp. 56-60, 2010.

[8] A. Y. Satpute, A. D. Bhoi, and T. Engineering, "ELECTRONIC SPEAKING SYSTEM FOR DUMB," vo!. 6, no. 3, pp. 1132-1139, 2013.

[9] M. Wald, "Captioning for Deaf and Hard of Hearing People by Editing Automatic Speech Recognition in Real Time", Proceedings of 10th International Conference on Computers Helping People with Special Needs ICCHP 2006, LNCS 4061, pp. 683- 690.

[10] R. R. Itkarkar and A. V. Nandi, "Hand gesture to speech conversion using Matlab," in 2013 Fourth International Conference on Computing, Communications and Networking Technologies (ICCCNT), 2013, pp. 1-4.

[11] Jingdong Zhao, Li Jiang, Shicai Shi, Hegao Cai, Hong Liu, G. Hirzinger, "A Five-fingered Underactuated Prosthetic Hand System",Proceedings of the 2006

[12] Harish, N. and Poonguzhali, S., 2015, May. Design and development of hand gesture recognition system for speech impaired people. In 2015 International Conference on Industrial Instrumentation and Control (ICIC) (pp. 1129-1133). IEEE.

[13] Vishal, D., Aishwarya, H.M., Nishkala, K., Royan, B.T. and Ramesh, T.K., 2017, December. Sign Language to Speech Conversion. In 2017 IEEE International Conference on Computational Intelligence and Computing Research (ICCIC)(pp. 1-4). IEEE.

[14] Vijayalakshmi, P. and Aarthi, M., 2016, April. Sign language to speech conversion. In 2016 International Conference on Recent Trends in Information Technology (ICRTIT) (pp. 1-6). IEEE.

[15] Heera, S.Y., Murthy, M.K., Sravanti, V.S. and Salvi, S., 2017, February. Talking hands—An Indian sign language to speech translating gloves. In 2017 International conference on innovative mechanisms for industry applications (ICIMIA) (pp. 746-751). IEEE.

[16] Bhaskaran, K.A., Nair, A.G., Ram, K.D., Ananthanarayanan, K. and Vardhan, H.N., 2016, December. Smart gloves for hand gesture recognition: Sign language to speech conversion system. In 2016 International Conference on Robotics and Automation for Humanitarian Applications (RAHA) (pp. 1- 6). IEEE.

[17] Jean, C., and Peter, B., “Recognition of Arm Gestures Using Multiple Orientation Sensors: Gesture Classification”, IEEE Intelligent transportation systems conference on electronics, Vol. 13, No. 1, pp. 334-520,2004.

[18]. Otiniano, R., and Amara, C. “Finger spelling recognition from rgb-d information using kernel descriptor”, IEEE Transactions on neural systems and rehabilitation engineering, Vol.28, No. 8, pp. 124-184, 2006.

[19]. Zhengmao, Z., Prashan, R., Monaragala, N., Malin, P. “Dynamic hand gesture recognition system using moment invariants”, IEEE Transactions on

neural networking and computing, Vol. 21, No. 1, pp. 1034- 1320, 2010.

[20]. Recognition of sign language gestures using neural networks by Peter Vamplew. Department of Computer Science, University of Tasmania [2]. Fels and G Hinton (1993), Glove-Talk: A Neural Network Interface Between a Data-Glove and a Speech Synthesiser, IEEE Transactions on Neural Networks, 4, 1, pp. 2-8 [3]. Johnston (1989), Auslan: The Sign Lang