DHWANI- THE VOICE OF DEAF AND MUTE

Amarthya Vishnu V[1], Amrita Sree S[2], Blessy Geevarghese[3], Jessline Elezabeth Jacob[4] , Er. Gokulnath G [5]

[1,2,3,4] B. Tech Student, Dept. of Computer Science and Engineering, Saintgits College of Engineering, Kerala, India
[5] Assistant Professor, Dept. of Computer Science and Engineering, Saintgits College of Engineering, Kerala, India

Abstract - Human beings with the ability to see, hear, and speak are one of God's most magnificent creations. However, some people do not have access to this precious benefit. Deaf-mute people are most affected since, speech is the primary medium of communication that allows individuals to exchange information, ideas, and expressions in both verbal and nonverbal ways. A hand gesture recognition system for deaf persons is used to communicate their thoughts to others. Folks have a hard time to understand what these people are saying since they utilize sign languages to connect with the rest of the world. As a result, this initiative makes use of gestures that are an integral part of such people's lives and converts them into words. The communication happens when others are able to understand and respond to one's message. This method bridges the communication gap between deaf-mute people and the general public, allowing for a more productive dialogue. This artificially intelligent system uses Kera's as a platform for converting gestures taken in real-time via camera and trained using a convolutional neural network (CNN) are converted into text as output.

Key Words: **CNN, Hand- Gesture, ReLU, deaf-mute**

CHAPTER 1

INTRODUCTION

Kitty O'Neil was an American stunt woman and speed racer, most famous in the 1970s. She was deaf and mute. But she was known as 'the fastest woman in the world. She conquered the world without hearing and speaking. She was truly an inspiration for many of the deaf and mute people. But in this current scenario, since almost everything is going online, people with speech and hearing disabilities are not able to communicate normally and convey their thoughts to others. Sign language is used by deaf and hard hearing people to exchange information between their own community and with other people. So, in our system, we lend a helping hand for the exceptional creation of God by providing them a platform on when they can communicate with others normally without any difficulty. There are some applications such as hand gesture technology which can detect sign language and convert it to human language. But these applications need to be installed in the system which they are using, which is not very easy. Server creating a uniformly usable plugin that can be used in any of the applications with the help of Google. It is much easier and less costly. The scope of this sign language detector is to expand to other applications such as zoom, teams, etc.

1.1 GENERAL BACKGROUND

The project is about helping the deaf and mute people in the country to work easily on the system, with the help of gestures. It includes static as well as dynamic gestures for the working of this system. Here machine learning, deep learning and image processing concepts are incorporated. The fundamental point of building a sign language acknowledgment is to make a characteristic collaboration among humans and computer. Motions can be static (pose or certain posture) which require less computational intricacy or dynamic (grouping of stances) which are more complicated however reasonable for constant conditions.

1.2 RELEVANCE

People use gestures like nodding and waving throughout their daily lives without even noticing it. It has become a major feature of human communication. New approaches of human computer interaction (HCI) have been developed in recent years. Some of them are based on human-machine connection via the hand, head, facial expressions, voice, and touch, and many are still being researched. The problem occurs when sign language is studied by self-learning method. Having the ability to sense gesture-based communication is an intriguing computer vision issue, but it's also incredibly useful for deaf-mute individuals to communicate with people who don't understand sign language. This project is mostly focused on deaf and mute people, who find it difficult to communicate with others. Since the signs presented by these persons are not seen by regular people, the meaning of motions will be displayed as text and this text will be transformed into dialogue. This allows deaf and mute people to communicate much more successfully. The proposed system provides a motion acknowledgement, which differs from commonly used frameworks in that it is obviously difficult for the visibly unable to deal with. Here, users can work the framework with the help of motions while remaining in a safe distance from the computer and without the need to use the mouse.

1.3 Socio-Economic Importance

The project is mainly focused on a socially relevant issue for our society's deaf and mute people. The project aims to make it easier for the deaf and mute community to interact with others by capturing motions with the help of a webcam and displaying the resulting message on the screen, which then generates text, making it much easier for them to connect with others.

1.4 Applications

- This system is used in webinars for interacting with deaf and mute people.
- This system is used to train mute people.
- This system can be used in live classes for interacting with deaf and mute people.

1.5 Advantages

- This project deals with providing easiness to the dumb to communicate with people using gestures which will be captured by the webcam and the resultant message being displayed on the screen, which gives it much more easiness for the deaf-mute people to communicate with others.
- This system is a plugin. So, there is no need of an external application.
- It needs only less storage space.

CHAPTER 2

LITERATURE SURVEY AND EXISTING SYSTEMS

In this chapter, the literature review and existing systems related to our project will be covered

2.1 LITERATURE SURVEY

Sign Language Recognition using 3D convolutional neural networks [1]. Jie Huang University of Science and Technology of China, Hefei, China; Wengang Zhou; Houqiang Li; Weiping Li. The goal of Sign Language Recognition (SLR) is to translate sign language into text or speech in order to improve communication between deaf-mute people and the general public. Due to the complexity and significant variances in hand gestures, this activity has a wide social influence, but it is still a difficult work. Existing SLR methods rely on hand-crafted characteristics to describe sign language motion and create classification models based on them. However, designing reliable features that adapt to the wide range of hand movements is difficult. To address this issue, a new 3D convolutional neural network (CNN) that automatically extracts discriminative spatial-temporal characteristics from raw video streams without any prior knowledge, reducing the need to build features are offered. Multi-channels of video streams, including colour information, depth clues, and body joint locations, are used as input to the 3D CNN to integrate colour, depth, and trajectory information in order to improve performance. On a real dataset acquired with Microsoft Kinect, the proposed model is validated and is shown that it outperforms the previous approaches based on hand-crafted features.

Real-Time Sign Language Recognition Using a Consumer Depth Camera [2]. Alina Kuznetsova; Laura Leal-Taixe; Bodo Rosenhan. In the field of computer vision and human-computer interaction, gesture detection remains a difficult job (HCI). A decade ago, the challenge appeared to be nearly impossible to complete using only data from a single RGB camera. There are additional data sources available due to recent improvements in sensing technologies, such as time-of-flight and structured light cameras, which make hand gesture detection more practical. It offers a highly precise approach for recognizing static gestures from depth data provided by one of the sensors specified above in this paper. Rotation, translation, and scale invariant characteristics are derived from depth pictures. After that, a multi-layered random forest (MLRF) is trained to categorize the feature vectors, resulting in hand sign recognition. When compared to a simple random forest with equal precision, MLRF requires substantially less training time and memory. This makes it possible to repeat the MLRF training method with minimal effort. To demonstrate the benefits of this method, it is tested on synthetic data, a publicly accessible dataset containing 24 American Sign Language (ASL) signals, and a new dataset gathered with the recently released Intel Creative Gesture Camera.

Indian sign language recognition using SVM [3]. Machine Vision Lab CSIRCEERI, Pilani, India J. L. Raheja; School of Instrumentation D.A.V.V. Indore, Pilani, India A. Mishra; Researcher, Pilani, India A. Chaudhary. People are always inspired to create new methods to engage with machines by needs and new technologies. This interaction can be for a specific purpose or as a framework that can be used in a variety of applications. Sign language recognition is a critical field in which ease of engagement with humans or machines will benefit many individuals. At the moment, India has 2.8 million people who are unable to talk or hear correctly. This research focuses on Indian sign identification in real-time using dynamic hand gesture recognition techniques. The acquired video was transformed to HSV color space for pre-processing, and skin pixels were used for segmentation. Depth data was also used in parallel to provide more accurate results. HuMoments and motion trajectories were recovered from image frames, and gestures were classified using a Support Vector Machine. The system was tested using both a camera and an MS Kinect. This type of device might be useful for teaching and communicating with deaf people.

Sign Language Recognition Using Convolutional Neural Networks [4]. University College London-Lourdes Agapito; University of Lugano, Switzerland-Michael M. Bronstein; Technische Universität Dresden, Dresden, Germany-Carsten Rother. The Deaf population and the hearing majority have an undeniable communication challenge.

Automatic sign language recognition innovations are attempting to break down this communication barrier. A recognition system based on the Microsoft Kinect, convolutional neural networks (CNNs), and GPU acceleration is the main focus of the contribution. CNN's can automate the process of feature building rather than creating intricately handcrafted features. It has a high level of accuracy in recognizing 20 Italian gestures. With a cross-validation accuracy of 91.7 percent, the predictive model may generalize to users and environments that were not present during training. In the ChaLearn 2014 Looking at People gesture spotting competition, the model have received a mean Jaccard Index of 0.789.

Sign language recognition using wifi[5]. Ma, Y., Zhou, G., Wang, S., Zhao, H. and Jung, W., 2018. Signfi: Sign language recognition using wifi. Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies, 2(1), pp.1-21. SignFi is a WiFi-based gesture recognition system that is proposed. SignFi's input is Channel State Information (CSI) measured by WiFi packets, and the classification mechanism is a Convolutional Neural Network (CNN). Existing WiFi-based sign gesture detection algorithms have only been evaluated on 25 movements using hand and/or finger gestures. SignFi can recognise 276 sign gestures with excellent accuracy, including head, arm, hand, and finger gestures. SignFi collects CSI measurements in order to record wireless signal properties of sign gestures. Raw CSI data are pre-processed to reduce noise and retrieve CSI changes across subcarriers and sample times.For sign gesture categorization, preprocessed CSI measurements are supplied into a 9-layer CNN.

The CSI traces are collected and tested SignFi in the lab and at home. There are 8,280 gesture instances, 5,520 from the lab and 2,760 from home, for a total of 276 sign motions. SignFi's average recognition accuracy for 5-fold cross validation utilising CSI traces of one user is 98.01 percent, 98.91 percent, and 94.81 percent for the lab, home, and lab+home environments, respectively. In the lab, it also executes tests with CSI traces from five distinct users. For 7,500 instances of 150 sign gestures performed by 5 distinct users, SignFi's average recognition accuracy is 86.66 percent.

American Sign Language recognition using rf sensing [6]. Sevgi Z. Gurbuz; Ali Cafer Gurbuz; Evie A. Malaia; Darrin J. Griffin; Chris S. Crawford; Mohammad Mahbubur Rahman; Emre Kurtoglu; Ridvan Aksu; Trevor Macks; Robiulhossain Mdrafi;. Many human-computer interaction technologies are built for hearing people and rely on vocalised communication, therefore users of American Sign Language (ASL) in the Deaf community are unable to benefit from these improvements. While video or wearable gloves have made significant progress in ASL detection, the use of video in homes has raised privacy problems, and wearable gloves significantly restrict movement and encroach on daily life. Methods: The use of RF sensors in HCI applications for the Deaf community is proposed in this paper. Regardless of lighting conditions, a multi-frequency RF sensor network is employed to provide non-invasive, non-contact measurements of ASL signing. Time frequency analysis with the Short-Time Fourier Transform reveals the distinctive patterns of motion contained in the RF data due to the micro-Doppler effect. Machine learning is used to examine the linguistic features of RF ASL data (ML). The information content of ASL signing is proven to be greater than that of other upper body actions seen in daily life, as evaluated by fractal complexity. This can be utilized to distinguish daily activities from signing, while RF data shows that non-signers' imitation signing is 99 percent distinguishable from native ASL signing. The classification of 20 native ASL signs is 72.5 percent accurate thanks to feature-level integration of RF sensor network data. Implications: RF sensing can be utilized to investigate the dynamic linguistic features of ASL and to create Deaf-centric smart environments for non-invasive, remote ASL identification. ASL data that is natural, not imitation, should be used to test machine learning techniques.

CHAPTER 3

OBJECTIVES AND PROPOSED INNOVATION

In this chapter, objectives and proposed innovation of our project will be discussed.

3.1 OBJECTIVES

In a number of computer applications, gestures offer an innovative interaction paradigm. A deaf and mute person's thoughts are communicated to others via a sign language recognition system. The majority of such individual interact with the outside world through sign language, which makes it difficult for others to comprehend. Deaf and mute people are the most affected since speaking is the most common way of communication. The goal of this project is to employ a plugin that allows users to interpret deaf and mute people's sign language.

3.2Broad Objectives

- Make a system allowing the deaf and mute to recognize hand gestures.
- The primary goal is to transform sign languages into text.

3.3Specific Objectives

- Gesture train design: -The camera records each gesture, which is subsequently taught.
- Feature extraction: -Applied to get characteristics that will aid in picture classification and recognition.
- Gesture recognition: -Every gesture made is recognized.
- Gesture to text translation: -Display the result on the screen when the motions have been recognized.

CHAPTER 4

HYPOTHESIS, DESIGN AND METHODOLOGY

In this chapter the hypothesis, design and methodology about the project will be covered.

4.1 HYPOTHESIS

Due to the challenges of vision-based problems such as varying lighting conditions, complex backgrounds, and skin color detection, efficient hand tracking and segmentation is the key to success in any gesture recognition. Variation in human skin color variations required the robust development of an algorithm for the natural interface. For object detection, color is a highly useful descriptor

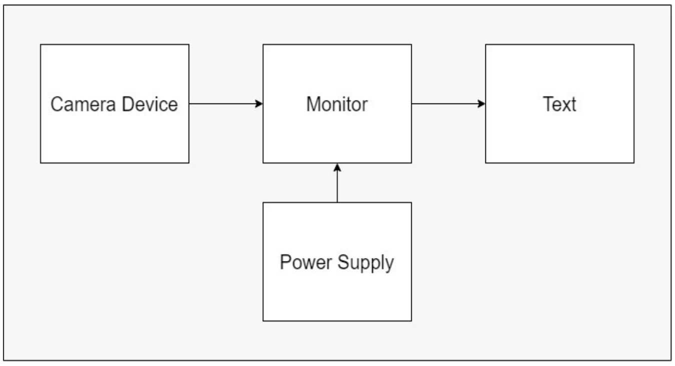

Fig 4.1.1 Architecture

4.2 DESIGN

The system is divided into 3 parts:

- Dataset Creation
- Gesture Training
- Gesture-to-Text

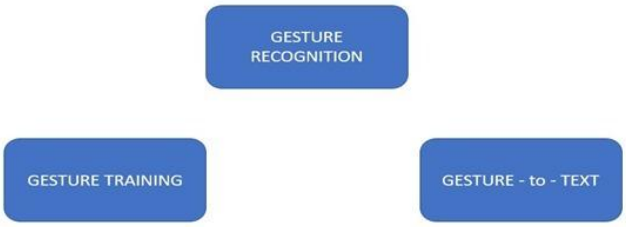

Fig 4.2.1 Design

A. Dataset Creation

This model will be having a live feed from the video camera and every frame that detects a hand in the ROI (region of interest) created will be saved in a directory (here gesture directory) that contains two folders train and test, each of the 10 folders contains images captured using Sign_Language.py The creation of the dataset is done by using a live camera feed with the help of OpenCV and creating an ROI that is nothing but the part of the frame where it needs to detect the hand-in for the gestures. The red box is the ROI and this window is for getting the live camera feed from the webcam. To distinguish between the background and the foreground, the background's cumulative weighted average is computed and removed from the frames that have some item in front of the background that may be recognized as foreground. Calculating the accumulated weight for certain frames (here for 60 frames) and the accumulated average for the backdrop accomplishes this.

B. Gesture Training

It deals with training the gestures collected from the deaf-mute individuals. The system trains the captured gestures using Convolutional Neural Network (CNN). Convolutional neural networks (CNN) have brought a revolution in the computer vision area. It also plays an important role for generic feature extraction such as scene classification, object detection, semantic segmentation, image retrieval, and image caption.

The convolution neural network contains three types of layers:

- Convolution layers: The convolution layer is the core building block of the CNN. It carries the main portion of the network’s computational load.
- Pooling layers: The pooling layer replaces the output of the network at certain locations by deriving a summary statistic of the nearby outputs. This helps in reducing the spatial size of the representation, which decreases the required amount of computation and weights.

Full connection layers: Neurons in this layer have full connectivity with all neurons in the preceding and succeeding layer as seen in regular Fourier Convolutional Neural Network. This is why it can be computed as usual by a matrix multiplication followed by a bias effect.

The system employed thirty-two convolutional filters of size 3x3 over a 200x200 picture with an activation function ReLu, followed by Max-Pooling. The Rectified Linear Unit, or ReLu, is a linear activation function with a threshold of 0, which means the model will take less time to train or execute. The goal of employing Max-Pooling is to minimize the geographic dimensions of the activation maps as well as the number of parameters employed in the network, lowering computational complexity. Following that, a dropout of 0.5 is employed to prevent the model from overfitting and to generate some picture noise augmentation.

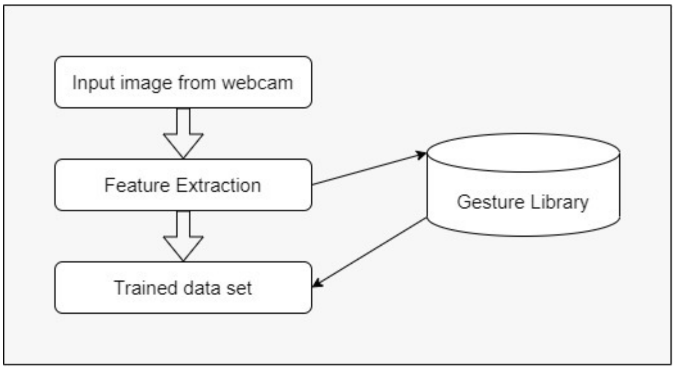

Fig 4.2.2 Gesture Training

C. Gesture-to-Text

This phase is concerned with identifying motions and turning them to text. Cropping the ROI detects the hand, which is then compared to the movements in the lexicon. Using an offline python package, the identified motions are translated to text. The training gestures are later recognized at this phase. The movements are exhibited inside a region of interest (ROI) that has been clipped. If the gesture shown by the user is present in the database, then the relevant text message is shown on the screen.

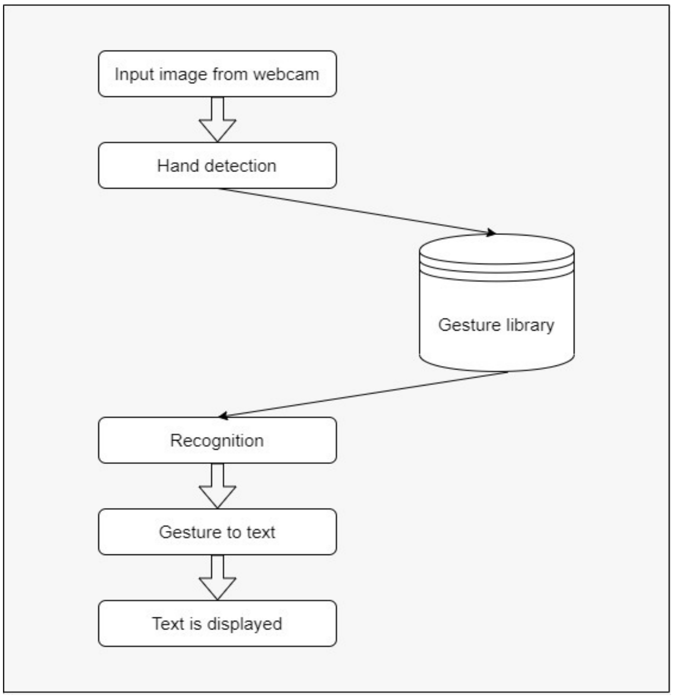

Fig 4.2.3 Gesture to Text-Flow chart

4.3 METHODOLOGY

A. DATA COLLECTION: It is a process of capturing image using web camera.

B. IMAGE PROCESSING: The pictures delivered are in RGB shading spaces. The hand motion turns out to be harder to section dependent on skin tone. Therefore, we convert the pictures to HSV shading space. HSV is an integral asset for further developing picture security by isolating splendor from chromaticity.

C. FEATURE EXTRACTION: After we have extricated the significant highlights, we can utilize include extraction to decrease the information. It likewise assists with keeping the classifier's precision and works on its intricacy. By lessening the picture size to 64 pixels, we had the option to get an adequate number of elements to viably order the American Sign Language signals.

D. CLASSIFICATION: Subsequent to dissecting the info hand picture. The motion grouping strategy is utilized to perceive the signal and afterward create the voice message in a like manner.

E. EVALUATION: The exhibition of the framework is evaluated subsequent to preparing.

F. PREDICTION: Predicts contribution of client and showcases result.

G. REAL TIME DETECTION: A continuous motion that is shown by the client is recognized by the framework.

H. DISPLAYING WORD SEQUENCE: The communication through signing is changed over to word and is shown in the chatbot.

CHAPTER 5

FUTURE SCOPE

The future scope of our project includes:

- **Augmented Reality**

  The rear future is in the hands of Augmented Reality. People in public places can communicate using sign language easily with others by the integration of augmented reality to this. So the sign language the person is showing will be displayed in the space as words so communication will be easier without the usage of a system.

- **Conversion of voice and text to sign language**

  In future work, proposed system's image processing part should be improved so that System would be capable of converting normal language to sign language. We will try to recognize signs which include motion. Moreover, we will focus on converting the sequence of gestures into word and sentences and then converting it into the speech which can be heard.

- **As a Translator**

  As we all use Google translator for translating normal languages, with the help of artificial intelligence this plugin can be modified and used as a translator for translating different types of sign languages.

- **Music**

  People with speech and hearing disability can sing song using this sign language system. There will be certain signs for musical notes. The person can show this particular musical note sign and the system will recognize the musical notes and will combine this in a specific pattern and will be played correspondingly using Artificial intelligence and Machine learning techniques with the help of computer vision.

CHAPTER 6

CONCLUSION

This system can be used to assist deaf-mute people to convey their messages to normal people without the assistance of an interpreter. In this undertaking work, communication through signing will be more useful for the simplicity of correspondence between quiet individuals and ordinary individuals. The undertaking basically targets lessening the hole of correspondence between quiet individuals and ordinary individuals. Here the strategy catches the quiet signs into discourse. In this framework, it beats the troubles looked by quiet individuals and helps them in further developing their way. The projected framework is extremely simple to convey to any place when contrasted with existing frameworks. To help the quiet individuals, the language gets changed over into text kind and on the advanced showcase screen, it will be shown. Who can't speak with ordinary individuals for example tragically challenged individuals the framework is especially useful. The essential element of the task is the one that will be applied in like manner puts that the recognizer of the signals might be an independent framework.

REFERENCES

[1] Kuznetsova, A., Leal-Taixé, L. and Rosenhahn, B., 2013. Real-time sign language recognition using a consumer depth camera. In Proceedings of the IEEE international conference on computer vision workshops (pp. 83-90).

[2] Kuznetsova, A., Leal-Taixé, L. and Rosenhahn, B., 2013. Real-time sign language recognition using a consumer depth camera. In Proceedings of the IEEE international conference on computer vision workshops (pp. 83-90).

[3] Raheja, J.L., Mishra, A. and Chaudhary, A., 2016. Indian sign language recognition using SVM. Pattern Recognition and Image Analysis, 26(2), pp.434-441.

[4] Pigou, L., Dieleman, S., Kindermans, P.J. and Schrauwen, B., 2014, September. Sign language recognition using convolutional neural networks. In European Conference on Computer Vision (pp. 572-578). Springer, Cham.

[5] Ma, Y., Zhou, G., Wang, S., Zhao, H. and Jung, W., 2018. Signfi: Sign language recognition using wifi. Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies, 2(1), pp.1-21.

[6] Gurbuz, S.Z., Gurbuz, A.C., Malaia, E.A., Griffin, D.J., Crawford, C.S., Rahman, M.M., Kurtoglu, E., Aksu, R., Macks, T. and Mdrafi, R., 2020. American sign language recognition using rf sensing. IEEE Sensors Journal, 21(3), pp.3763-3775.