REVIEW ON HAND GESTURE TRANSLATORS

Lavish Kumar[*1], Nikhil Sinha[*2], Raghav Gupta[*3], Ravi Prakash[*4], Amrita Bhatnagar[*5]

[*1,2,3,4]Student, Department of Computer Science and Engineering, Inderprastha Engineering College, Ghaziabad, India.

[*5]Asst. Professor, Department of Computer Science and Engineering, Inderprastha Engineering College, Ghaziabad, India.

ABSTRACT

The motive of our work is to connect individuals with deaf and mute people and provide a platform where they can interact with them. According to a survey conducted in 2017, only 17% of normal people know sign language. To solve this problem, we either need to learn sign language or we need a platform that can help us to interact with them. We will be using deep learning methods to solve this problem and develop a model that will help deaf and mute people.

Keywords: Machine Learning, Computer Vision, Deep Learning, Sign Language, Image Processing.

I.   INTRODUCTION

There are roughly 7,00,000 to 9,00,000 deaf people in the world. But there is not a single platform that can help in interacting with them. In a survey conducted there are very less people who can use sign language. We need to overcome this barrier. To do so two ways are either to learn Sign Language or build a platform that can help in interaction. The first way is certainly impossible to achieve so the only option left is to build a platform. To do that we will be using Deep Learning methods along with Image Processing. We will be using the Long Short Term Memory(LSTM) and Sequential model for training of data that we will get using the Image Processing Method. For image processing, we will be using OpenCV which is an image and video-related library. To get and distinguish between gestures we will be using MediaPipe which is a tool offered by Google for motion detection. The motions we get from MediaPipe will be converted into a matrix in the form of float values and further used for the training of the model. The final Deep Learning model that we will receive would be able to predict trained gestures and give desired output in understandable language.

II.   LITERATURE REVIEW

Hand Gesture Recognition is a difficult task that takes time and effort to complete. There appears to be a slew of existing or forthcoming options. Physical device-based solutions are the most common. Supervised learning and classification algorithms are commonly used in Artificial Intelligence-based solutions. These solutions provide good accuracy but there is still room for improvement. These solutions are either costly or inconvenient for the users. One of its example can is using a third person to translate hand gestures into other non-gesture based language. To overcome this many solutions are introduced.

Murakami and Taguchi [1], in 1991 published the first article on sign language recognition. They devised a system for detecting sign language motions that makes use of neural networks and computer vision techniques. The research provided by them opened many doors for this field. Many researchers started working in this field. Later Laptev and Lindeberg [2] devised a method for tracking multistate hand models in 2001. Particle filtering and the hierarchy of multiscale image characteristics are two techniques they utilized to accomplish this. This helped the future researchers to derive a method that can be used to detect and trace hand from the frame of images. To determine the area of hand gestures, they used various hand positions, orientations, sizes, and postures. Laptev, Linderberg, and Bretzner [3] continued their work in the field of hand gesture identification later in 2002. They introduced algorithms for representing hand postures in multiscale color image features at various sizes in multiscale color image characteristics. The performance of the previous model improved as a result of this.

Later, Arpit Mittal [4] proposed an OpenCV-based approach for detecting objects. OpenCV is a library that provides functions to deal with camera. In 2002, S. A. Mehdi and Y. N. Khan [5] published a paradigm for hand gesture recognition that utilized gloves. They employed two sensors to track hand orientation and five sensors for each finger. They employed seven sensors in all and reached an accuracy of around 87 percent. The sensor

based gesture became quite popular and many other researchers proposed their study on this topic. Following that, a slew of models and approaches for predicting sign language emerged. Wang and Popovic's model [6], which they presented in 2009, was one of them. This model employed colored gloves to monitor real-time hand motions and classified them using the K-Nearest Neighbors method. However, Tharwat et al. [7], Baranwal et al. [8], Rekha et al. [9], and Kurdyumov et al. [10] exhibited better results in contrast to this using Support Vector Mechanism approach. SVM became popular as the prediction become easy and it does not require sensor based gloves.

Tsitsoulis and Bourbakis [11] published a study in 2013 in which they explained how to recognize hands from a single picture. They said that single frame or a group of frames can be used to detect sign language. From here many machine learning based models emerged that made perdition on multiple frames. In 2021, Arpita and Akshit [12] proposed a study using MediaPipe and various other algorithms to recognize hand motions in a continuous set of frames. Using the Support Vector Mechanism for American Sign Language, they achieve a 99.15 percent accuracy rate. Souradeep [13] also used MediaPipe in conjunction with a Recurrent Neural Network to anticipate five motions. He employed the Long Short Term Memory model and achieved a 94 percent accuracy rate. As a result, we may infer that hand gesture identification is a difficult problem that necessitates a large amount of data and a complicated approach. To forecast American Sign Language, we will use MediaPipe and the Long Short Term Memory model. To construct our dataset for our project, we'll use MediaPipe, NumPy, Pandas, and OpenCV.

III. PROPOSED METHODOLOGY

To promote awareness about deaf and mute people and to set a bridge so that people can easily communicate with them, our work will be mainly focused on developing a medium that will make this happen. For handling the frontend, we will be using Flask framework which is an open-source python framework for web development, also we will be using Python, NumPy, OpenCV, TensorFlow, and MediaPipe for backend and main execution. People will use their webcam to interact with our model. The input taken will be converted into a set of frames which are later used for the prediction of the sign associated with it. For doing so, we will be using LSTM and Sequential model along with MediaPipe. To access the camera, we will be using the OpenCV library. MediaPipe is the tool provided by Google which will help in getting co-ordinate points of hand and posture. We will use the LSTM and Sequential model for training the model. The coordinates extracted will be processed with the help of NumPy and then trained using Deep Learning Model.

IV. CONCLUSION

We can conclude that there exist multiple solutions that uses sensor gloves but the most convenient solution uses machine learning algorithm to classify the gestures. Most successful models are based on deep learning techniques that train a model on image frames or group of image frames. Than a classification algorithm is used to train that model and predict the hand gesture. Hence sensor based solutions might have introduced new aspects for hand gesture prediction but most accurate solutions are achieved by machine learning and deep learning algorithm. Among the machine learning algorithm Support Vector Machine have performed better than other classification algorithm.

V. ACKNOWLEDGEMENT

VI. REFERENCES

[1] Murakami K, Taguchi H. 1991. Gesture recognition using recurrent neural networks. In: Proceedings of the ACM SIGCHI conference on Human factors in computing systems, pp 237–242. https

[2] ://dl.acm.org/doi/pdf/10.1145/10884 4.10890 0

[3] I. Laptev, T. Lindeberg. 2001. Tracking of Multi-state Hand Models Using Particle Filtering and a Hierarchy ofMulti-Scale Image Features. DOI:10.1007/3-540-47778-0_6

[4]   L. Bretzner, I. Laptev, T. Lindeberg. 2002. Hand gesture recognition using multi-scale colour features, hierarchical models and particle filtering. DOI:10.1109/AFGR.2002.1004190

[5]   Arpit Mittal, Andrew Zisserman, P. Torr. 2011. Hand detection using multiple proposals. DOI:10.5244/C.25.75

[6]   S. A. Mehdi and Y. N. Khan, "Sign language recognition using sensor gloves," In the proceedings of the 9th International Conference on Neural Information Processing, 2002, ICONIP '02., Vol.5, pp. 2204-2206, 2002,

[7]   DOI: 10.1109/ICONIP.2002.1201884

[8]   Wang RY, Popović J. 2009. Real-time hand-tracking with a color glove. ACM Trans Graph 28(3):63

[9]   Tharwat A, Gaber T, Hassanien AE, Shahin MK, Refaat B. 2015. Sift-based arabic sign language recognition system. In: Springer Afro-European conference for industrial advancement, pp 359–370. DOI: 10.1007/978-3- 319-13572-4_30

[10]   Baranwal N, Nandi GC. 2017. An efficient gesture based humanoid learning using wavelet descriptor and MFCCtechniques. Int J Mach Learn Cybern. 8(4):1369–1388

[11]   Rekha J, Bhattacharya J, Majumder S. 2011. Hand gesture recognition for sign language: a new hybrid approach.

[12]   In: International Conference on Image Processing, Computer Vision, and Pattern Recognition (IPCV), pp 80–86

[13]   Ruslan Kurdyumov, Phillip Ho, Justin Ng. 2011. Sign Language Classification Using Webcam Images. DOI:10.1.1.374.9029

[14]   A. Tsitsoulis, N. Bourbakis. 2013. Towards Automatic Hands Detection in Single Images. DOI:10.1007/978-3-642-41184-7_48

[15]   Arpita Halder, Akshit Tayade. 2021. Real-time Vernacular Sign Language Recognition using MediaPipe and Machine Learning. International Journal of Research Publication and Reviews Vol.2, Issue.5, Page 9-17

[16]   Souradeep Ghosh. 2021. Proposal of a Real-time American Sign Language Detector using MediaPipe andRecurrent Neural Network. IJCSE Vol.9, Issue.7