Grenze International Journal of Engineering and Technology, June Issue

Hand Gesture Recognition for Hearing Impaired and Non-Verbal Individuals using Mediapipe

Mrs. Sadhana B[1], Mythri B[2], Aneesha D Souza[3], Renuka Durgadas Naik[4] and Anusha Rao[5]
[1]Assistant Professor, Information Science and Engineering Canara Engineering College, Mangalore
Email: sadhana.ise@gmail.com,
[2-5]Student, Information Science and Engineering Canara Engineering College, Mangalore
Email: mythribb@gmail.com, aneeshadsouza2002@gmail.com, renunaik4556@gmail.com, anusharao878@gmail.com

***Abstract*— People with hearing impairments and those who are non-verbal frequently struggle to express themselves and comprehend others in a world that is largely based on verbal communication. Technology has become a potent ally in bridging this communication gap. this paper introduces a novel method for hand gesture recognition with the goal of empowering non-verbal and hearing-impaired people by giving them an easy-to-use means of communication. Hand gesture recognition is a crucial research area in human-computer interface due to its flexibility and user-friendliness. It is used to communicate information among disabled individuals or control devices. However, challenges include illumination variation, nonuniform backgrounds, hand size and shape differences, and high interclass similarities between hand gesture poses, making it challenging to develop an efficient technique. Google's MediaPipe framework is a solution that enables real-time hand gesture recognition for hearing impaired and non-verbal individuals. This system uses computer vision and machine learning to interpret hand gestures and convert them into meaningful actions or messages. The system identifies a wide range of common hand gestures, which can be mapped to specific actions or words. The system can be implemented on various devices, making it accessible and versatile. This research highlights the potential of MediaPipe for enhancing the lives of hearing impaired and non-verbal individuals, bridging the communication gap, fostering inclusivity, and improving their overall quality of life. This technology has the potential to completely change how these people interact with their environment by utilizing real-time hand tracking and gesture recognition capabilities for enabling meaningful communication process.**

***Index Terms*— Hand Gesture Detection, MediaPipe, OpenCV, TensorFlow, Real-Time Processing.**

I. INTRODUCTION

The aim of this project is to create an application-specific computer vision based system that can recognize and comprehend hand gestures. The system employs a camera to capture live video data, which is then processed in real-time using machine learning and computer vision algorithms to identify and classify hand gestures. Individuals with hearing impairments or those who are nonverbal confront distinct obstacles in expressing themselves and understanding others in a world that mainly relies on verbal communication. Traditional techniques of communication, such as sign language or written text, can be useful but are not always easy or accessible. With

Grenze ID: 01.GIJET.10.2.267_1

technological advancements, notably in the field of computer vision, a viable solution in the shape of Hand Gesture Recognition has emerged. This technology enables hearing-impaired and nonverbal individuals to communicate more effectively through gestures, creating new opportunities for interaction and inclusion. Google's open-source framework MediaPipe has evolved as a versatile and user-friendly solution for hand gesture identification. It provides a collection of pre-trained models and tools for a variety of computer vision tasks, such as hand tracking and gesture detection. The framework is user-friendly and adaptable, making it an excellent choice for designing applications that address the specific needs of hearing challenged and nonverbal users. In conclusion, MediaPipe hand gesture recognition is a breakthrough technology with enormous potential to improve the lives of hearing impaired and nonverbal individuals. It fosters inclusivity, freedom, and accessibility in many facets of life, in addition to providing a means of communication. As technology advances, we may expect increasingly more complex and user-friendly programs that will allow these people to express themselves and interact with the world around them. With the continuous development and use of technologies like as MediaPipe as given in paper is Sign to letter translator system using a hand glove proposed by Mr.H. El Hayek,, which is getting closer to a future in which communication boundaries are broken down and everyone has the ability to fully engage in society. A Hand Recognition model uses a webcam to capture hand images, segment and analyze them using MediaPipe's landmark estimation, detects similar gestures, uses text-to-speech, calculates FPS, and uses a confusion matrix for key point classification with 95% accuracy.
The organization of the paper is formatted with Literature Survey, Methodology, Results and Conclusion

II. Literature Survey

Mr. Dixit, V., &Mr. Agrawal, A proposed a triphasic approach consisting of Hand Detection & Tracking, Feature Extraction, and Gesture Recognition. Initially application of hand tracking and detection algorithms to capture hand motion and extract positions. Trajectory-based features are derived from the hand movements and a Hidden Markov Model (HMM) is designed for each gesture. HMM is used to model the generative sequences. This method is tested on a dataset containing 16 gestures resulting in an impressive average recognition rate of 91%. Thus outperforming traditional approaches like PCA, ANN, SVM, and DTW.
Mr.Gadekallu, Dr.T. R., Alazab, M., Mr.Kaluri, R.,Mr. Maddikunta, P. K. R., Mr.Bhattacharya, S.,Mr. Lakshmanna, K., & Mrs.Parimala, M proposed a way to detect hand gestures using Crow-Search-Based Convolutional Neural Networks (CSA-CNN) prototype in the Human-Computer Interaction (HCI) domain. They utilized a Kaggle hand gesture dataset and implemented one-hot encoding along with CSA to optimize the model's hyper-parameters. This eliminated irrelevant parameters and resulted in a 100 percent accuracy in both training and testing phase. Thus this model has a better approach than conventional approaches in enhancing accuracy for classifying hand gestures.
Pro.Praneeth, Mr.G., Naveen, R.,Mr. Prakash, A. S.,Mrs. Rashmi, M., &Mr. Guddeti, R. M. R proposed a hand gesture related medium for regulating computer behaviour using deep learning and personal dataset. Mediapipe Python framework provides handy solutions for the users and PyAutoGUI is a module used to direct mouse functions through program. The model was constructed for Support Vector Machine (SVM), Decision Trees, KNN, and a DNN having three layers. It was demonstrated after plenty of experimental findings that DNN model has the greatest accuracy.
Mr.Sen, A., Mr. Mishra, T. K., & Dash, R proposed an ensemble approach to extenuate issues associated with conventional convolutional neural network (CNN) architectures. Starting off with gesture portion detection through a binary thresholding-based background separation method. Following, which the extraction of contours and hand segmention is implemented. The scores obtained from these models are meaned to make an optimized ensemble model for the ultimate prediction. The system is then evaluated on two publicly available datasets and one self-constructed dataset. Thus, this ensemble model demonstrates superior performance collated to existing progressive approaches.
Pro.Marc Marschark and Dr.Patricia Elizabeth Spencer proposed that Sign language connects people with impaired hearing or verbal abilities through gestures, facial expressions, and symbols, enabling communication between the deaf and the dumb. Hand gesture recognition is challenging due to segmentation. This paper presents a novel algorithm for hand gesture recognition using RGB color spaces and models, addressing the challenges of segmentation in hand language communication for hearing and speech-impaired individuals.
Mr.Vijay P. K.,Dr. Suhas, N. N.,Prof. Chandrashekhar, C. S. &Mr. Dhananjay, D. K proposed the study developed a generic solution for hand gesture recognition using machine learning algorithms. The prototype, developed using C++ and OpenCV, integrates posture and gesture classification with any human-computer interface. It can process

new hand features, learn different postures, and dynamic gestures.This work is a small contribution towards achieving the results needed in the field of hand gesture recognition, as it is a challenging task compared to traditional devices.

Mr.H. El Hayek,Mr. J. Nacouzi, Pro.A. Kassem, Mr.M. Hamad and Mr.S. El-Murr proposed a hand glove system that can convert sign language motions into letters shown on an LCD screen is one example of recent research employing sensors. Another example is the sign to text translation system using a hand glove. Deaf and mute persons can use this approach to communicate with those who don't know sign language. Flex resistors, a microprocessor, and an LCD display make up the system. It aims to overcome communication barriers and ensure effective message transmission.

Pro.Camillo Lugaresi, Pro.Jiuqiang Tang, Mr.Hadon Nash,Mr. Chris Mc-Clanahan,Mrs. Esha Uboweja,Mr. Michael Hays, Mr.Fan Zhang, Mr.Chuo-Ling Chang, Mr.Ming Guang Yong,Mr. Juhyun Lee, Mr.Wan-Teh Chang, Mr.Wei Hua,Mr. Manfred Georg, and Mr. Matthias Grundmann proposed A real-time on-device hand tracking solution predicts human hand skeletons from a single RGB camera for AR/VR applications. Implemented via MediaPipe, it offers real-time inference speed on mobile GPUs and high prediction quality. MediaPipe Hands is an end-to-end hand tracking solution that predicts 2.5D landmarks without specialized hardware, enabling easy deployment to commodity devices for gesture control and AR/VR applications.

Mr.Akoum, Alhussain, and Mr.Nour Al Mawla Gesture recognition is a nonverbal or nonvocal communication method that uses movement in the human body to express a specific message. Hand gestures, ranging from basic hand motions to more intricate ones, can be used in interpersonal interactions. Today, various frameworks and machine learning libraries have been developed to facilitate the development of AI apps for users, including MediaPipe, which includes features like Face recognition, Face Mesh, Iris, Hands, Pose, Holistic, Hair segmentation, Object recognition, Box Tracking, Instant Motion Tracking, Objection, and KIFT.

Pro.Ustunug A, Mr.Cevikcan proposed that hand gesture recognition is crucial in Human-Computer Interactions (HCI) development technology. HCI allows computers to recognize, understand, and execute hand gestures as commands without requiring physical contact with devices. One solution for a hand gesture detection system is the MediaPipe framework, which is used in this study to create a basic application for user guides. User guides provide step-by-step instructions for managing a specific system and reduce user irritation by enabling them to recognize, comprehend, and decipher technical issues.

Pro.Harris, M., & Mr.Agoes, A. S proposes the development of Tongue-Operated Assistive Technology (TDS) for people with severe disabilities to control computers and their environment using their tongue motion. TDS translates users' volitional tongue movements into commands, which are then interfaced to external devices. The new interface allows the external TDS headset to communicate with the smartphone and relay sensor data to the signal processing algorithm, turning raw data into control commands. Other proposed technologies include sarcastic gadget technology, smart gloves, Qureshi deaf-mute communication interpreters, switching rate changes, and interdisciplinary design of assistive technology for the third world.

III. METHODOLOGY

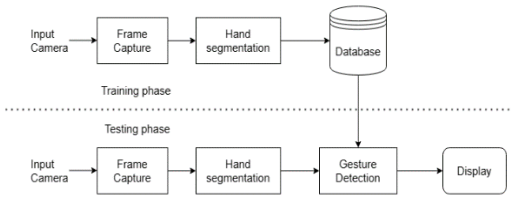

Fig. 1: Methodology diagram for Hand gesture recognition model

Fig.1 represents very simply the operation of the Hand gesture recognition model. Firstly, a functional webcam is a must as the image will be processed. Thus, to the camera we provide the object to be estimated which are the hands. The frames captured contain the images of the hand. The captured image is horizontally flipped for Mirror display so that the person has a natural view of the display. These images are subject to Hand segmentation. When a person shows the hand gestures to the camera the background is also unintentionally included. Hand segmentation is the process of identifying and focusing on the areas in the image frames which relate to the hands of the person only.

MediaPipe has many convenient solutions out of which we particularly use the Hands solution. Because Hands solution contains the landmark estimation for hands. Landmarks mean the joints which are numbered so that the

gestures can be easily implemented. But we include a Model logging feature which allows use of a label for the corresponding image, stored in the form of co-ordinates. This will be our custom dataset. Another data shall contain the corresponding hand gestures to the labels provided. Causing an indirect relation between co-ordinates logged in and hand gestures though labels. In our model we have included 9 random gestures which represent the custom dataset. However, we have the liberty to customize any gesture we want into the model with the use of the Model logging feature. This concludes the training phase of the model.
Split the custom dataset so that the testing phase is 30% of the custom dataset using sklearn Python library. When a person provides the hand gesture to the input camera the model detects the gesture most similar to the one in the dataset which has the co-ordinates logged in. Then the result is displayed in Real-Time for the user. This can be visualized using boundary box and drawing over the image using MediaPipe. used a custom function for drawing visual elements on the image which are represent the key points (landmarks) with circles and lines which connect the landmarks present.
Also included a text-to-speech function to vocalise the gestures made. The model checks for changes in the identified hand gesture and vocalizes the gesture using TTS only once for each new gesture. The custom function to calculate the Frames per second (FPS). The purpose of the class is to calculate the corresponding FPS number by measuring the amount of time that passes between consecutive frames. The number of prior frame time differences to save and average is represented by the buffer length, which the constructor initializes when creating the class. To calculate the system tick count and frequency which are utilized to calculate the millisecond interval between frames—the class makes use of the OpenCV library (cv2). The get method is in charge of determining the frames per second (FPS) by measuring the difference in time between the current and previous frames, adding that difference to the internal deque, and averaging the time differences. The outcome is a rounded frame-per-second (FPS) figure that offers a more representative and seamless assessment of the frame rate throughout a given buffer duration. In order to maintain fluid and real-time visual feedback in applications like video processing, object detection, and gesture recognition, it is necessary to calculate frames per second (FPS). This allows for performance monitoring in computer vision applications.
The included Python script uses the MediaPipe module for hand tracking to construct a hand motion recognition system. The script records video from a camera, uses the MediaPipe Hands module to identify the user's hand, and then categorizes the hand motions according to their historical data and important points. In addition to supporting logging the data for machine learning model training, the system is capable of recognizing motions like pointing. To understand hand movements, the application combines point history classification models with keypoint classification models. The script also has capabilities like the ability to log gesture data for training reasons and to provide text-to-speech feedback for identified movements. It uses the pyttsx3 library for text-to-speech capabilities and the OpenCV library for image processing and visualization.
The code is divided into modular components for drawing landmarks, bounding rectangles, and managing various modes for data logging, and is arranged into methods like vocalize_text, get_args, and main.

IV. Results And Discussion

Here, we discuss the result obtained after training the model with custom dataset.

A. Results

The custom dataset is obtained after logging the key points. We have considered 9 hand gestures to be recognized by the application. Below is confusion matrix obtained for key point classification Figure 2.

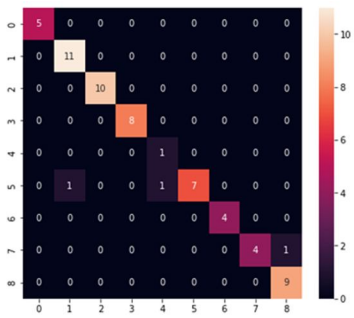

Fig. 2: Confusion matrix for the custom gestures.

The classification report containing the recall, precision and f1-score is estimated. The accuracy is calculated to be 95%. Fig.3 is the table containing details of classification report below:

Custom Gestures	**Classification Report**		
	precision	***recall***	***f1-score***
0 -thumbs up	1.00	1.00	1.00
1-peace	0.92	1.00	0.96
2-ok	1.00	1.00	1.00
3-rock	1.00	1.00	1.00
4-zero	0.50	1.00	0.67
5-one	1.00	0.78	0.88
6-three	1.00	1.00	1.00
7-four	1.00	0.80	0.89
8-open	0.90	1.00	0.95

Fig. 3: Table containing classification report for custom hand gestures

B. Discussion

The most common observation from the literature review is that Hand Gesture detection was usually done using CNN. CNN is used for image analysis and considers the pixels in extreme detail. Since we are more concerned about Real-Time processing, MediaPipe framework is the better option. Only need to estimate the gesture and the color etc. are irrelevant. Some papers have considered the application of MediaPipe however they have been for a specific dataset and the gestures are mostly hardcoded. Our model is flexible towards any type of gesture within a limit so customization is extremely favourable for the users. Our model includes key points logging which is different compared to the previous adaptations of Hand Gesture Recognition models.

V. Conclusion And Future Scope

A wide range of hand movements may be recognised by the system in real-time, enabling more logical and organic interactions between people and computers. The system can effectively classify new hand movements and adapt to a variety of applications and circumstances by utilising machine learning and computer vision methods. The proposed hand gesture system has a huge potential impact. It might transform the way sign language is recognised, allowing deaf people to interact with hearing people more successfully. It might also improve the usability and accessibility of different digital devices and platforms, giving users a more organic and intuitive way to connect with technology. We plan to work on dynamic hand gestures in the future to enhance this model.

References

[1] Pro.Marc Marschark and Dr.Patricia Elizabeth Spencer, “The Oxford Handbook of Deaf Studies”, Language, and Education, Vol. 1, 2nd ed. Oxford University Press, 2011

[2] Mr.Vijay, P. K.,Dr. Suhas, N. N.,Prof. Chandrashekhar, C. S. &Mr. DHANANJAY, D. K. “Recent Developments in Sign Language Recognition”: A Review. International Journal on Advanced Computer Engineering and Communication Technology, 2012

[3] Mr.H. El Hayek,Mr. J. Nacouzi, Pro.A. Kassem, Mr.M. Hamad and Mr.S. El-Murr."Sign to letter translator system using a hand glove," The Third International Conference on e-Technologies and Networks for Development, Beirut, Lebanon, 2014

[4] Mr.Akoum, Alhussain, and Mr.Nour Al Mawla,"Hand gesture recognition approach for ASL language using hand extraction algorithm." Journal of Software Engineering and Applications, 2015

[5] Mr. Dixit, V., &Mr. Agrawal, A,“ Real time hand detection & tracking for dynamic gesture recognition”. International Journal of Intelligent Systems and Applications, 2015.

[6] Pro.Ustunug A, Mr.Cevikcan, Industry 4.0: “Managing The Digital Transformation”, Springer Series in Advanced Manufacturing, Switzerland.,2018.

[7] Pro.Camillo Lugaresi, Pro.Jiuqiang Tang, Mr.Hadon Nash,Mr. Chris Mc-Clanahan,Mrs. Esha Uboweja,Mr. Michael Hays, Mr.Fan Zhang, Mr.Chuo-Ling Chang, Mr.Ming Guang Yong,Mr. Juhyun Lee, Mr.Wan-Teh Chang, Mr.Wei Hua,Mr. Manfred Georg, and Mr. Matthias Grundmann. “Mediapipe: A framework for building perception pipelines”, 2019

[8] Mr.Lugaresi, C.,Mr. Tang, J., Nash, H., Mr.McClanahan, C., Mr.Uboweja, E.,Mr. Hays, M., ... & Pro.Grundmann, M. “Mediapipe: A framework for building perception pipelines”,2019
[9] Mr.Zhang, F.,Mr. Bazarevsky, V.,Mr. Vakunov, A., Mr.Tkachenka, A., Sung, G., Mr.Chang, C. L., &Mr. Grundmann, M. “Mediapipe hands On-device real-time hand tracking”,2020
[10] Mr.Gadekallu, Dr.T. R., Alazab, M., Mr.Kaluri, R.,Mr. Maddikunta, P. K. R., Mr.Bhattacharya, S.,Mr. Lakshmanna, K., & Mrs.Parimala, M. “Hand gesture classification using a novel CNN-crow search algorithm. Complex Intell Syst”,2021
[11] Pro.Harris, M., & Mr.Agoes, A. S.Applying “hand gesture recognition for user guide application using MediaPipe”. In 2nd International Seminar of Science and Applied Technology Atlantis Press, 2021
[12] Pro.Praneeth, Mr.G., Naveen, R.,Mr. Prakash, A. S.,Mrs. Rashmi, M., &Mr. Guddeti, R. M. R. “Vision-based Hand Gesture Interface for Real-time Computer Operation Control”. In TENCON, 2022
[13] Mr.Sen, A., Mr. Mishra, T. K., & Dash, R. “A novel hand gesture detection and recognition system based on ensemble-based convolutional neural network”. Multimedia Tools and Applications, 2022