iJRASET
International Journal For Research in Applied Science and Engineering Technology



INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Volume: 11 **Issue:** XII **Month of publication:** December 2023

DOI: https://doi.org/10.22214/ijraset.2023.57310

www.ijraset.com

Call: 08813907089 | E-mail ID: ijraset@gmail.com

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.538
Volume 11 Issue XII Dec 2023- Available at www.ijraset.com

Sign Language Detection of English Alphabets for Deaf and Dumb People

Raghav Jaju[1], Parth Karwa[2]
Dept. of Electronics and Telecommunication, Vishwakarma Institute of Information Technology, Pune, India

Abstract: ***This research explores the implementation of a robust system for sign language detection designed to facilitate communication for individuals who are deaf and dumb. Leveraging the capabilities of Python, Mediapipe, OpenCV, and Scikit Learn, our proposed system focuses on real-time hand sign recognition of English alphabets. The framework employs Mediapipe for hand landmark detection, enabling precise tracking of hand gestures. OpenCV is utilized for image processing, allowing efficient handling of video streams. The combination of these tools enables the extraction of relevant features from hand signs, forming the basis for our recognition model. The core of our system is built upon the Random Forest algorithm from Scikit Learn, implementing machine learning for the classification of hand signs. This approach ensures adaptability to various hand shapes and orientations, contributing to the versatility of the system. The proposed solution aims to bridge communication gaps for individuals with hearing and speech impairments, empowering them to express themselves through sign language. Through real-time detection and interpretation of hand signs, our system provides a valuable tool for enhancing communication and fostering inclusivity for the deaf and dumb community.***

Keywords: Sign language detection, hand sign recognition, Python, Mediapipe, OpenCV, Scikit Learn, Random Forest.

I. INTRODUCTION

Sign language serves as a vital medium of communication for individuals with hearing and speech impairments, enabling them to express thoughts, ideas, and emotions.

However, the lack of automated tools for the real-time interpretation of sign language poses a significant challenge in fostering effective communication for the deaf and dumb community. In response to this challenge, our research endeavours to design and implement an innovative system, leveraging cutting-edge technologies such as Python, Mediapipe, OpenCV, and Scikit Learn, to enable the detection and interpretation of hand signs representing English alphabets.

Communication is a fundamental human right, and our mission is to bridge the existing communication gap for those who rely on sign language. The proposed system employs a multi-faceted approach to address the complexities inherent in recognizing the diverse and nuanced hand signs that form the basis of sign language.

Through the integration of advanced tools, we aim to create a real-time, efficient, and adaptive solution capable of accurately interpreting hand gestures.

Mediapipe, a powerful library for hand landmark detection, forms the cornerstone of our system. By precisely capturing the landmarks on the user's hand, we can derive valuable spatial information critical for accurate sign interpretation. Complementing this, OpenCV facilitates streamlined image processing, enabling the efficient handling of video streams and enhancing the overall robustness of the system.

The heart of our solution lies in the implementation of the Random Forest algorithm from Scikit Learn. This machine learning algorithm excels in classification tasks, making it an ideal choice for recognizing the diverse hand signs that comprise sign language. The adaptability of Random Forest to different hand shapes and orientations ensures that our system is versatile and capable of accommodating the unique signing styles of individuals.

The significance of our research extends beyond the realms of technology; it directly impacts the lives of individuals with hearing and speech impairments. By providing a reliable and instantaneous means of interpreting sign language, our system aims to empower the deaf and dumb community, offering them a tool to communicate effortlessly in a world that has traditionally posed communication barriers.

In this pursuit, our research aligns with the broader objectives of promoting inclusivity, accessibility, and equal opportunities for all. By harnessing the potential of technology, we aspire to contribute to a more inclusive society where every individual, regardless of their abilities, can communicate freely and express themselves without constraints.

©IJRASET: All Rights are Reserved

II. LITERATURE SURVEY

Valentin Bazarevsky, Yury Kartynnik, Andrey Vakunov et.al have presented BlazeFace, a lightweight and well-performing face detector tailored for mobile GPU inference. It runs at a speed of 200-1000+ FPS on flagship devices. This super-realtime performance enables it to be applied to any augmented reality pipeline that requires an accurate facial region of interest as an input for task-specific models, such as 2D/3D facial keypoint or geometry estimation, facial features or expression classification, and face region segmentation. Our contributions include a lightweight feature extraction network inspired by, but distinct from MobileNetV1/V2, a GPU-friendly anchor scheme modified from Single Shot MultiBox Detector (SSD), and an improved tie resolution strategy alternative to non-maximum suppression.[1]

Arpita Halder et.al have presented a model with an average accuracy of 99% in most of the sign language dataset using MediaPipe's technology and machine learning, our proposed methodology show that MediaPipe can be efficiently used as a tool to detect complex hand gesture precisely. Although, sign language modelling using image processing techniques has evolved over the past few years but methods are complex with a requirement of high computational power. Time consumption to train a model is also high. From that perspective, this work provides new insights into this problem. Less computing power and the adaptability to smart devices makes the model robust and cost-effective. Training and testing with various sign language datasets show this framework can be adapted effectively for any regional sign language dataset and maximum accuracy can be obtained. Faster real-time detection demonstrates the model's efficiency better than the present state-of-arts.[2]

Hira Ansar, Ahmad Jalal et.al have developed an efficient HGR system for healthcare muscle exercise via point-based and full-hand features and a reweighted genetic algorithm. Features proposed in this method include Euclidean distance, the cosine of angles (i.e., α, β, γ), area of irregular shapes, SONG mesh, and chain model to select the optimal features. GWO with RGA is used to optimize, train, and recognize different gestures for muscle exercise. Our proposed system outperformed other HGR systems in terms of accuracy at 92.1%, 93.1%, 88.2%, 90.8%, and 85.3% over the Sign Word, Dexter1, Dexter + Object, STB, and NYU datasets, respectively. Then, precision, recall, and F1 scores were also measured for overall gesture recognition in all datasets.[3]

Saud S. Alotaibi et.al have discussed a novel approach for hand gesture recognition is proposed by applying five different stages i.e. preprocessing, hand detection, landmark detection, feature extraction, and classification, using geometric features for character recognition. The proposed HGR system targets the domain of HCI, HRI, and VR. The performance of the proposed geometric features approach is evaluated on the MNIST dataset on static gestures, also system performance is tested on ASL dataset for system validation. Experimental results show that our system competes with other state-of-the-art methods.[4]

III. METHODOLOGY

A. Mediapipe

Mediapipe is an open-source cross-platform framework provided by Google to build a pipeline for processing perceptual data from different modalities such as video and audio. Using Google's MediaPipe framework, focusing on its advantages over Open Pose. MediaPipe employs regression analysis for finger coordinates, offering computational efficiency and improved accuracy with a smaller detection range. It extracts 21 3D coordinates from a monocular camera frame, normalizing X and Y by the bounding box. The Z coordinate indicates depth, crucial for proximity assessment. This approach ensures advanced hand and finger tracking with reduced processing demands, particularly suitable for low-performance environments like mobile platforms, showcasing MediaPipe's effectiveness in comparison to other models.

B. Data Collection

The data samples were collected through a dummy web interface developed using Java Script and HTML which was hosted on a local web server. The website platform supported functionality to capture images, process them and record the hand track points as results after feature extraction.

The images were captured using a standard Webcam integrated with a laptop. The samples were collected from 4 volunteers who performed the hand gestures to be captured by the webcam with their hands at a distance ranging between 1.5 feet to 3 feet from the webcam. This distance was sufficient enough to capture the full gesture and was a good estimate to mimic the distance between two people in a conversation in real life. The hand gestures corresponding to 26 of American Sign Language Alphabets and 10 numbers were performed, and for each sign 100 images were captured t. The backgrounds of the image while capturing the samples were kept a mix of plain background with only the hand in the frame and a natural background with the person performing the gesture in the frame.

©IJRASET: All Rights are Reserved | SJ Impact Factor 7.538 | ISRA Journal Impact Factor 7.894 |

C. Hand Landmark Model

After the palm detection is over the whole image our subsequent hand landmark model performs precise key point localization of 21 3D hand-knuckle coordinates inside the detected hand regions via regression, that is direct coordinate prediction. The model learns a consistent internal hand pose representation and our Hand landmark model is robust even to partially visible hands which re blurry sometimes and self-occlusions. To obtain ground truth data in our model, we have manually used 30K real-world images with 21 3D coordinates, as shown below (we take Z-value from image depth map, if it exists per corresponding coordinate). To cover the possible hand poses and also to provide additional supervision on the nature of hand geometry, we also render a high-quality synthetic hand model over various backgrounds and map it to the corresponding 3D coordinates.



Fig1. Hand Coordinates

D. Data

Our dataset is captured by ourself and consists of 3600 images, each gesture has a dataset of 100 images. The data consists of data of alphabets and numeric. Then the data is divided in the ratio of 80:20 for training and testing of the data where random forest is used.

1) Feature Extraction:

The captured images were passed to a Hand tracking model from the MediaPipe framework which was deployed using TensorFlow js. The Hand tracking model's parameters and weights were adjusted as per our experiment's requirements. The Hand Tracking model uses a Blaze Palm detector to detect the palm in the captured image and passes the section of the image with the palm in it to a Hand Landmark Model [14]. The segment for the feature extraction process returned a Tensor of float values which contained the coordinates of 21 Landmark points of a Hand in 3D space. These coordinates serve as the data values for further carrying out the task of American Sign Language recognition.

E. Architecture

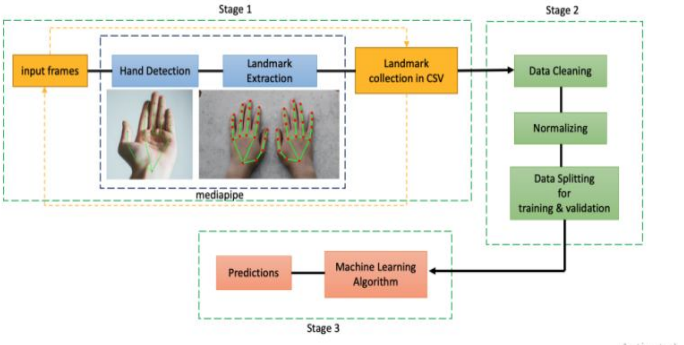


Fig2.Architecture of the project

A hand recognition system using MediaPipe and machine learning involves three main stages: input data processing, machine learning training, and real-time prediction and activation.

1) Stage 1: Input Data Processing:

a) Input Frames: The system receives a continuous stream of video frames as input.

b) Hand Detection: MediaPipe Hands is employed to detect the hand region within each frame.

c) Landmark Extraction: Key landmarks of the detected hand, such as fingertip positions and palm center, are extracted.

d) Data Collection: The extracted landmarks are stored in a CSV file for further processing.

2) *Stage 2: Machine Learning Training:*

a) *Normalization:* The collected landmark data is normalized to a common scale and range to ensure consistency.

b) *Data Splitting:* The normalized data is divided into training and validation sets.

c) *Machine Learning Algorithm:* A machine learning algorithm, such as a neural network, is trained on the training set to establish the relationship between hand landmarks and the desired output (e.g., hand gesture, sign language).

3) *Stage 3: Real-time Prediction and Activation*

a) *Predictions:* The trained machine learning algorithm is used to predict the desired output for the input hand landmarks.

b) *Activation:* The predicted output is translated into a corresponding action or control signal. For instance, a specific hand gesture could trigger a virtual object manipulation or a sign language translation.

This hand recognition system has potential applications in various domains, including sign language translation, gesture-based control of virtual environments, and augmented reality experiences. The system first detects the hand in each video frame using MediaPipe Hands. It then extracts the key landmarks of the hand and collects them in a CSV file. The system then normalizes the collected data and splits it into training and validation sets. The system trains a machine learning algorithm on the training set to learn the relationship between the hand landmarks and the desired output. Finally, the system uses the trained machine learning algorithm to predict the desired output for the input hand landmarks and activates it.

This system can be used for a variety of applications, such as sign language translation, gesture control of virtual objects, and augmented reality.

F. Algorithm Used

Random Forest Classifier uses a series of decision trees to label a sample. The performance of the algorithm depends upon the number of decision trees(n) which the algorithm forms to make a prediction. To find the optimal value of the number of decision trees the classifier was tested for different values of n in the range (1,200).

IV. RESULTS

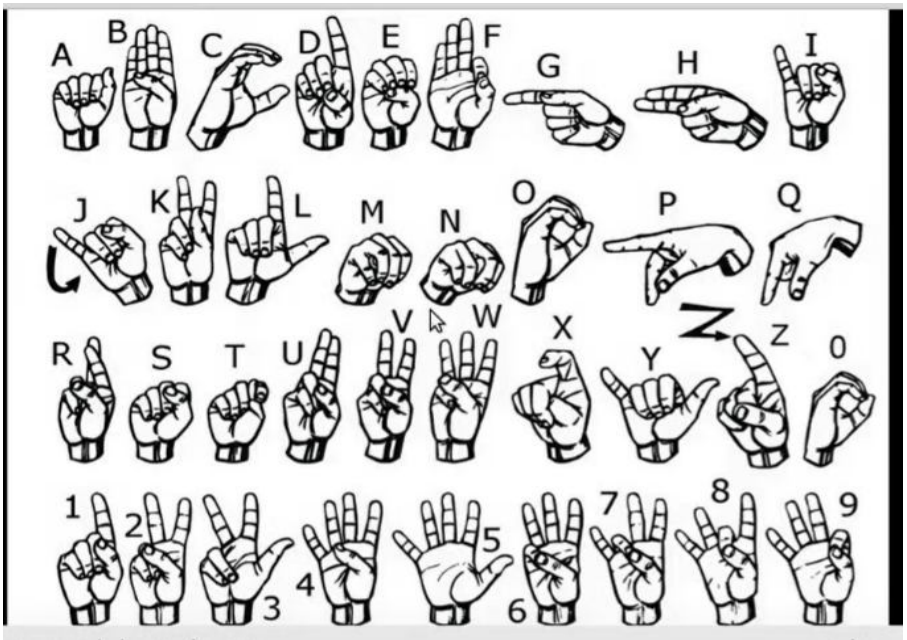


Fig3. Sign Template

This is the template for reference from where we have referred to the hand sign for our dataset and after this, we have applied the random forest where we have got an impressive result of an accuracy of 100 % because the dataset is small. Here are some of the results after the recognition

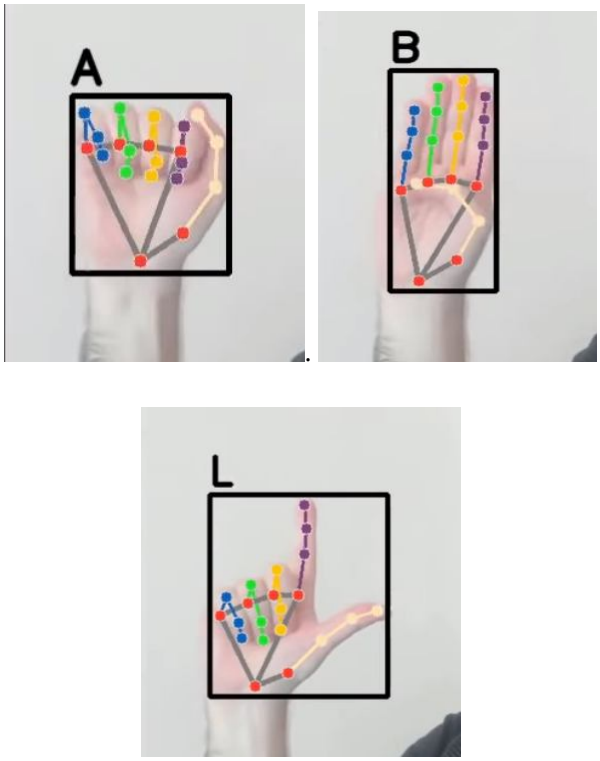


Fig4. Results

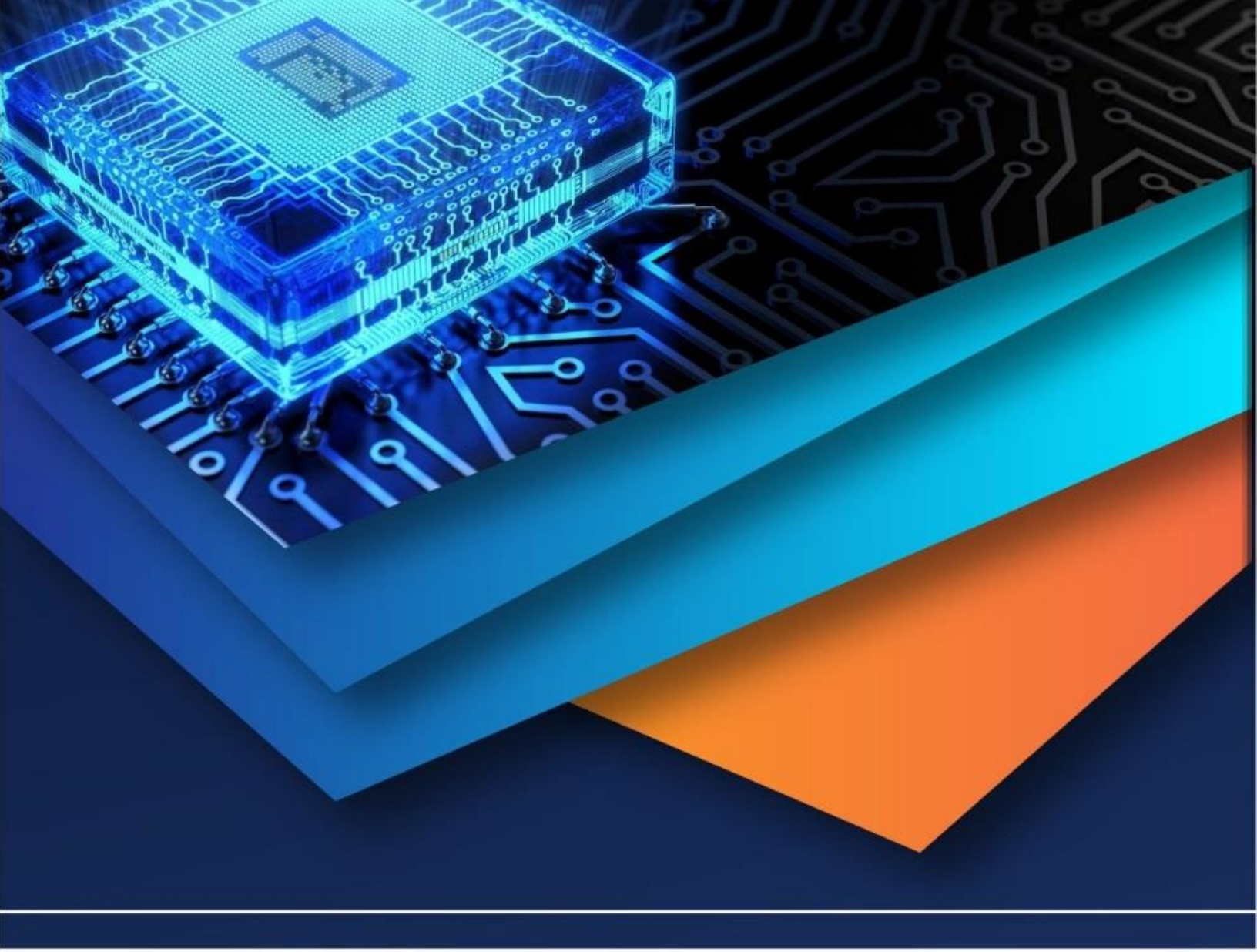
V. CONCLUSION

We have employed two types of bounding boxes, cuboidal and cubical, for hand gesture analysis. The cuboidal bounding box, the smallest cuboid encapsulating the 21 reference points, determines its dimensions from the minimum and maximum coordinate values for each axis. The cubical bounding box, the smallest cube fitting all reference points, adjusts its edge length uniformly based on the maximum difference among coordinate ranges. To normalize coordinates and facilitate generalization across gestures, two transformations are applied: shifting of origin and scaling. Shifting centers, the bounding box at the origin, while scaling ensures uniform sizing of all images to 255x255x255, enhancing classification algorithm efficiency. These transformations collectively establish a common reference frame, allowing effective comparison and pattern generalization for accurate gesture recognition.

REFERENCES

[1] Bazarevsky, Valentin, et al. "Blazeface: Sub-millisecond neural face detection on mobile gpus." arXiv preprint arXiv:1907.05047 (2019).
[2] Halder, Arpita, and Akshit Tayade. "Real-time vernacular sign language recognition using mediapipe and machine learning." Journal homepage: www. ijrpr. com ISSN 2582 (2021): 7421.
[3] Ansar, Hira, et al. "Hand gesture recognition based on auto-landmark localization and reweighted genetic algorithm for healthcare muscle activities." Sustainability 13.5 (2021): 2961.
[4] H. Ansar et al., "Hand Gesture Recognition for Characters Understanding Using Convex Hull Landmarks and Geometric Features," in IEEE Access, vol. 11, pp. 82065-82078, 2023, doi: 10.1109/ACCESS.2023.3300712.

©IJRASET: All Rights are Reserved | SJ Impact Factor 7.538 | ISRA Journal Impact Factor 7.894 |



10.22214/IJRASET

45.98

IMPACT FACTOR: 7.129

IMPACT FACTOR: 7.429

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Call : 08813907089 (24*7 Support on Whatsapp)