Realsense/camera based system for tracking Hand Pose

J.V.S Kalyan Aashish
M22AI561
Indian Institute of Technology
Jodhpur

Rakesh Reddy Kadapala
M22AI608
Indian Institute of Technology
Jodhpur

Krishna Kumari Ravuri
M22AI567
Indian Institute of Technology
Jodhpur

***Abstract*—This proposal centers on the development of an interactive sign language learning tool for differently-abled children within the context of Industry 4.0 and Human-Computer Interactions (HCI). The core objective is to harness gesture recognition technology to facilitate effective sign language acquisition through a user-friendly application. A laptop with an integrated webcam is employed for cost-efficiency and accessibility.**

The project leans on the MediaPipe framework, integrating machine learning to create a robust sign language recognition system. This technology enables computers to comprehend sign language gestures, enabling communication without physical device contact.

The envisioned application functions as an interactive sign language learning companion, providing step-by-step guidance and support for differently-abled children. It addresses identification, comprehension, and common learning challenges, empowering users to navigate the intricacies of sign language independently.

The experimental approach involves real-time image capture through the webcam, offering a readily available and budget-friendly solution. A diverse range of sign language gestures is trained and recognized within the system, each representing a unique facet of communication. These recognized gestures facilitate language learning, allowing users to track progress and access resources based on their produced sign gestures.

Our proposal advocates for the integration of sign language recognition via MediaPipe into our application, ensuring a comprehensive and engaging learning experience for differently-abled children. The transformation of conventional manual language learning into an interactive, user-friendly tool is paramount. This project aims to empower children with varying abilities to learn sign language interactively, equipping them with a vital means of communication and fostering inclusivity in education.

I. Introduction

Introduction:

The motivation behind this project is rooted in the recognition of the challenges faced by differently-abled children in acquiring sign language skills, a crucial means of communication for many. Traditional learning methods often lack the engagement and accessibility necessary to make the learning process effective and enjoyable for these children. We aim to address this disparity by providing an interactive and user-friendly solution that leverages the readily available technology of laptops equipped with webcams.

Our project's primary objectives are to develop a robust and responsive sign language recognition system within the MediaPipe framework and to create an engaging, step-by-step sign language learning companion application. This application will serve as a comprehensive guide for differently-abled children, offering them the means to identify, understand, and navigate the intricacies of sign language independently.

The contributions of this project extend beyond the development of a technical solution. By providing differently-abled children with an interactive and engaging tool for sign language acquisition, we aim to enhance their overall communication abilities, boost their confidence, and foster inclusivity in education. We believe this project can significantly impact the lives of these young learners by equipping them with vital communication skills.

In the following sections, we will delve into the technical details of our project, including the methodology, implementation, and expected outcomes.Through this project, we aspire to contribute to the broader goal of making education more accessible and inclusive for differently-abled individuals.

II. Planned Work:

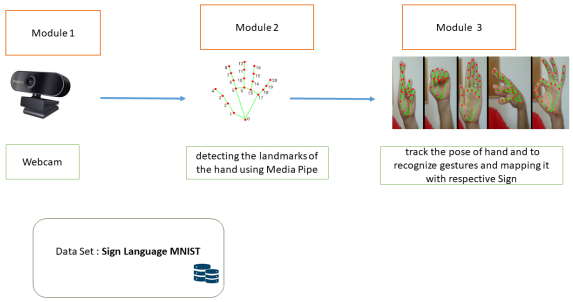

Fig. 1. Our Approach

A. Webcam Module:

This module functions as an optical sensor, capturing real-time video streams of hand gestures by acquiring individual video frames, which are used for further analysis.

B. Landmark Detection Module (MediaPipe):

This phase processes each video frame using the MediaPipe framework's Landmark Detection Module, identifying and cataloging significant hand landmarks. The result is a sequential dataset of 2D coordinates representing the spatial configuration of these landmarks over time.

C. Gesture Recognition Module:

This module interprets hand gestures over time, mapping observed hand poses to specific sign gestures based on predefined associations between landmark positions and sign language symbols. It also analyzes the temporal evolution of landmark positions to distinguish between different signs.

- Gesture Mapping: This step translates hand configurations into sign language symbols based on predefined associations.
- Temporal Analysis: This aspect focuses on recognizing temporal patterns in sign language gestures.

D. Sign Language MNIST Dataset:

This dataset serves for training and evaluating our gesture recognition model, comprising annotated hand gesture images corresponding to American Sign Language alphabet letters (excluding J and Z), labeled from 0 to 25.

- Training: During training, machine learning or deep learning models learn to recognize and interpret hand gestures.
- Testing: The trained model makes real-time predictions based on input from the Landmark Detection Module, allowing evaluation against the Sign Language MNIST dataset for accuracy assessment.

In summary, our sign language recognition system relies on the MediaPipe Landmark Detection Module for real-time hand gesture analysis, complemented by a Gesture Recognition Module and the Sign Language MNIST Dataset for training and evaluation. This system aims to enhance communication for individuals who primarily use sign language.

III. Resources

A. Hardware

Computer, Computer Webcam

B. Dataset

Most probably the follwing datasets will be used.

1.Sign Language MNIST Dataset
https://www.kaggle.com/datasets/datamunge/sign-language-mnist
2.Hand Gesture Recognition Database
https://www.kaggle.com/datasets/gti-upm/leapgestrecog
3.STATIC HAND GESTURE ASL DATASET
https://ieee-dataport.org/open-access/static-hand-gesture-asl-dataset

IV. References:

1.Boby RA, Prakash R, Saha SK, Matsumaru T, Sharma P, Jaitly S. Calibration and statistical techniques for building an interactive screen for learning of alphabets by children. International Journal of Advanced Robotic Systems. 2017;14(3). doi:10.1177/1729881417703939

2.International Journal of Advanced Robotic Systems, 14(3):1729881417703939. R. Damindarov, C. A. Fam, R. A. Boby, M. Fahim, A. Klimchik and T. Matsumaru, "A depth camera-based system to enable touch-less interaction using hand gestures," 2021 International Conference "Nonlinearity, Information and Robotics" (NIR), Innopolis, Russian Federation, 2021, pp. 1-7, doi: 10.1109/NIR52917.2021.9666090.

3.Thwe, Phyu Thè, May. (2019). Hand Gesture Detection and Recognition System: A Critical Review. International Journal of Computer (IJC). 32. 64-72.

4.Hand landmarks detection guide — MediaPipe https://developers.google.com/mediapipe/solutions/guide

5.Jyotishman Bora, Saine Dehingia, Abhijit Boruah, Anuraag Anuj Chetia, Dikhit Gogoi, Real-time Assamese Sign Language Recognition using MediaPipe and Deep Learning, Procedia Computer Science, Volume 218, 2023, Pages 1384-1393, ISSN 1877-0509, (https://www.sciencedirect.com/science/article/pii/S1877050923001175) https://doi.org/10.1016/j.procs.2023.01.117.