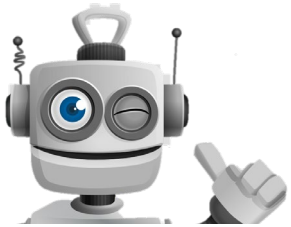
Computer Vision: Project Phase 2

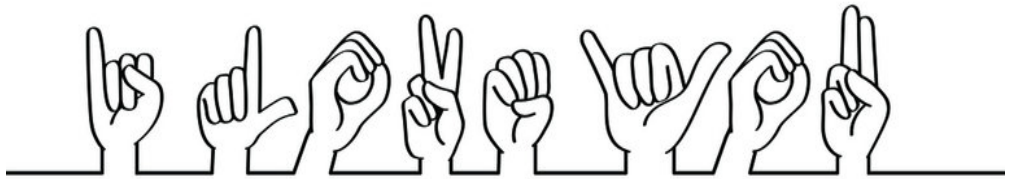
RUTGERS
THE STATE UNIVERSITY OF NEW JERSEY



Subject Code: 16:198:534:01
Subject: Computer Vision
Course Instructor: Prof Ahmed Elgammal

Contents

Usecase: Sign Language Detection



Name: Harish Udhaya Kumar
NetID: Hu33

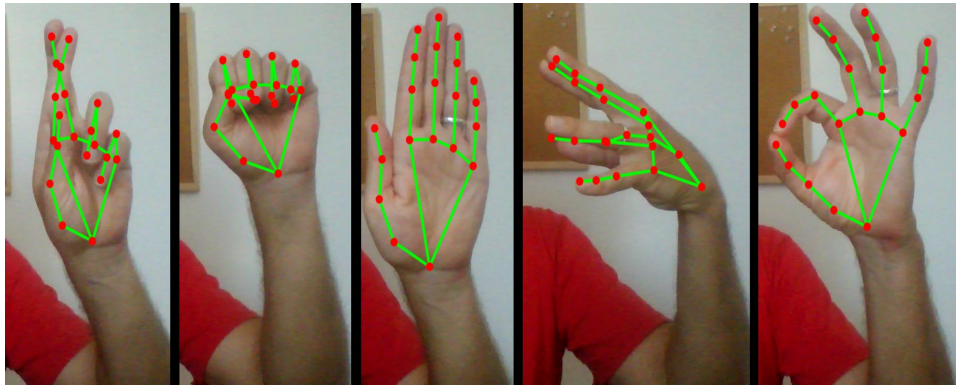
Activity Recognition

RUTGERS
THE STATE UNIVERSITY
OF NEW JERSEY

ACTIVITY RECOGNITION USING LSTM AND MEDIA PIPE
This project showcases how to utilize the Media pipe library over LSTM neural network to have high efficiency for activity recognition.

USE CASE: Sign Language Detection
In a vision-based approach, the features corresponding to the palms, finger position and joint angles are estimated, which are then used to perform recognition. This method requires acquiring images or videos of the signs through a camera and processing using image processing techniques.

Studying ASL promotes better awareness of and sensitivity to the deaf and hard of hearing community. As someone proficient in ASL, you will develop a strong appreciation for deaf culture, and you can promote understanding and acceptance of the language, among others.



Why LSTM?

LSTM networks are **well-suited to classifying, processing and making predictions based on time series data**, since there can be lags of unknown duration between important events in a time series

LSTMs are developed to deal with the vanishing gradient problem that can be encountered when training traditional RNNs

They have more & more controlling knobs, which control the flow and mixing of Inputs as per trained Weights. And thus, bringing in more flexibility in controlling the outputs

LSTMs provide us with a large range of parameters such as learning rates, and input and output biases no need for fine adjustments

Why not traditional feature extraction/edge detection using CNN?

MediaPipe

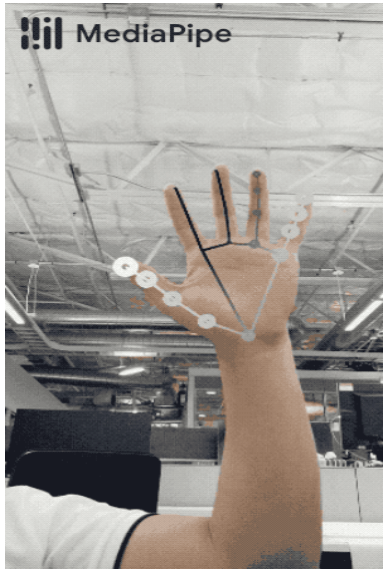
MediaPipe offers cross-platform, customizable ML solutions for live and streaming media.

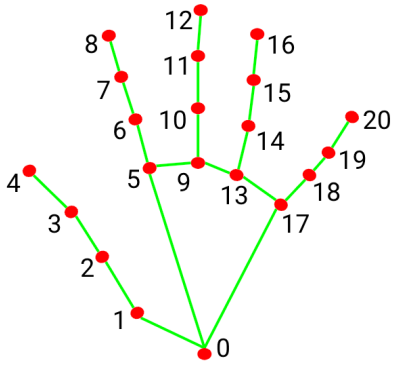
The ability to perceive the shape and motion of hands can be a vital component in improving the user experience across a variety of technological domains and platforms.

MediaPipe Hands is a high-fidelity hand and finger tracking solution. It employs machine learning (ML) to infer 21 3D landmarks of a hand from just a single frame.

MediaPipe Hands is a high-fidelity hand and finger tracking solution. It employs machine learning (ML) to infer 21 3D landmarks of a hand from just a single frame.

For example, it can form the basis for sign language understanding and hand gesture control, and can also enable the overlay of digital content and information on top of the physical world in augmented reality. While coming naturally to people, robust real-time hand perception is a decidedly challenging computer vision task, as hands often occlude themselves or each other (e.g. finger/palm occlusions and hand shakes) and lack high contrast patterns.



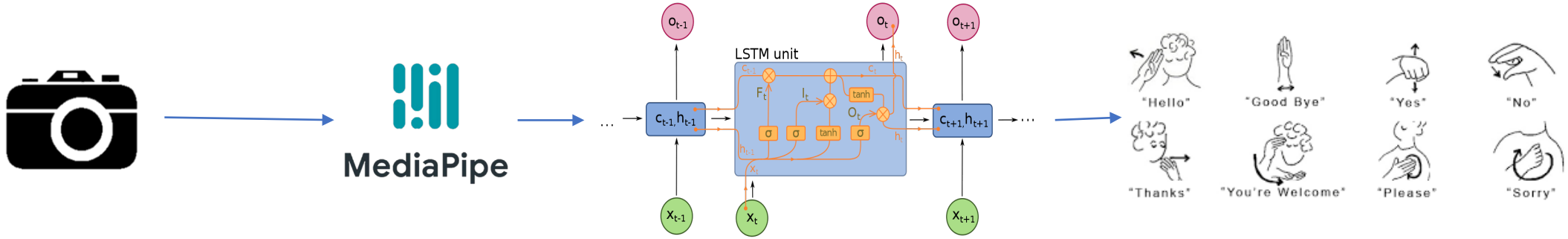


0. WRIST
1. THUMB_CMC
2. THUMB_MCP
3. THUMB_IP
4. THUMB_TIP
5. INDEX_FINGER_MCP
6. INDEX_FINGER_PIP
7. INDEX_FINGER_DIP
8. INDEX_FINGER_TIP
9. MIDDLE_FINGER_MCP
10. MIDDLE_FINGER_PIP
11. MIDDLE_FINGER_DIP
12. MIDDLE_FINGER_TIP
13. RING_FINGER_MCP
14. RING_FINGER_PIP
15. RING_FINGER_DIP
16. RING_FINGER_TIP
17. PINKY_MCP
18. PINKY_PIP
19. PINKY_DIP
20. PINKY_TIP

Approach

- Utilize Camera to capture the live streaming visual input.
- Run Mediapipe library to capture the features of the human hand
- Input this data into LSTM network for training and testing the model.

A normal CNN uses 3.5 million parameters in neural network to perform sign language recognition. However, using LSTM and media pipe, we will be able to do the same task with approximately half a million parameters.



Camera → Mediapipe → LSTM Neural network → Sign language output

Biblography

RUTGERS
THE STATE UNIVERSITY
OF NEW JERSEY

[1]https://www.who.int/news-room/fact-sheets/d etail/deafness-and-hearing-loss

[2] S.Masood, A.Srivastava, H.Thuwal, M.Ahmad. “Real-Time Sign Language Gesture (Word) Recognition from Video Sequences Using CNN and RNN”

[3] A. S. Elons ; Howida A. Shedeed ; M. F. Tolba ”Arabic sign language recognition with 3D convolutional neural networks” 2017

[4] 2021 Joint International Conference on Digital Arts, Media and Technology with ECTI Northern Section Conference on Electrical, Electronics, Computer and Telecommunication Engineering

[5] https://deepai.org/publication/activity-recognition-for-autism-diagnosis