International Research Journal of Engineering and Technology (IRJET) e-ISSN: 2395-0056
Volume: 09 Issue: 06 | June 2022 www.irjet.net p-ISSN: 2395-0072

Hand Gesture Recognition System Using Holistic Mediapipe

Ritesh Kate[1], Pranav Brahmabhatt[2], Prof. Swati Dhopte[3], Prof. Tushar Mane[4]

[1,2] *Student, Department of Computer Science Engineering MIT School of Engineering, Pune, Maharashtra, India*
[3,4] *Professor, Department of Computer Science Engineering MIT School of Engineering, Pune, Maharashtra, India*

Abstract - *Sign languages are visual languages produced by the movement of the hands, face, and body. In this paper, we evaluate representations based on skeleton poses, as these are explainable, person-independent, privacy-preserving, low-dimensional representations. Basically, skeletal representations generalize over an individual's appearance and background, allowing us to focus on the recognition of motion. But how much information is lost by the skeletal representation? We perform two independent studies using two state-of-the-art pose estimation systems. We analyze the applicability of the pose estimation systems to sign language recognition by evaluating the failure cases of the recognition models. Importantly, this allows us to characterize the current limitations of skeletal pose estimation approaches in sign language recognition.*

Key Words: *Holistic Mediapipe, Sign language recognition [SLR], LSTM, Computer Vision, Deep Learning, RNN.*

1. INTRODUCTION

Deaf people communicate via hand signs, regular folks have a hard time understanding what they're saying. As a result, systems that recognize various signs and deliver information to ordinary people are required. To create a virtual tracking system for people in need, which will be accomplished through image processing and human action recognition. This is primarily for persons who are unable to communicate with others. Models consisting of several CNN layers followed by numerous LSTM layers are often used to predict Real-Time Sign Language. The precision of these cutting-edge models, on the other hand, is quite low. This approach, Mediapipe Holistic with LSTM Model, on the other hand, provides significantly higher accuracy. This method yielded better outcomes with a smaller amount of data. This model trained much faster because it used less parameters, resulting in a shorter calculation time.

1.1 Motivation and Background

Sign Language Recognition strives to develop algorithms and methods for accurately identifying the sequences of symbols produced and understanding their meaning. Many SLR methods treat the problem as Gesture Recognition (GR). Therefore, research has so far focused on identifying the positive characteristics and methods of differentiation in order to properly label a given signal from a set of possible indicators. Sign language, however, is more than just a collection of well-articulated gestures.

1.2 What is Real Time Gesture Recognition?

Gesture Detection detection is an active research field in Human-Computer Interaction technology. It has many applications for visual environment control and sign language translation, robot control, or music creation. In this machine learning project on hand touch recognition, we will create a Real-Time Touch Viewer using the MediaPipe and Tensorflow framework in OpenCV and Python. many types of different tech and computer vision algorithms to translate the signal language. many types of tech and computer vision algorithms to translate sign language.

1.3 Sign Language Recognition System

It is always difficult to communicate with someone who has a hearing problem. People with hearing and speech problems are now able to communicate their feelings and thoughts to the whole world in sign language, which has become an indelible solution. It simplifies and simplifies the process of integration between them and others. However, simply developing sign language is not enough. This blessing comes with many strings attached. For someone who has never read or known a different language, the movement of symbols is often mixed and confused. However, with the advent of more automated signaling methods, this long-standing communication gap can now be closed. We provide a sign language recognition program based on Acquisition of Personal Action in this paper. In this study, the user should be able to take hand-drawn pictures using a webcam and the system will predict and display the name of the captured image. The main goal of this project is to help the deaf, dumb, and blind community to communicate as an ordinary person using this project.

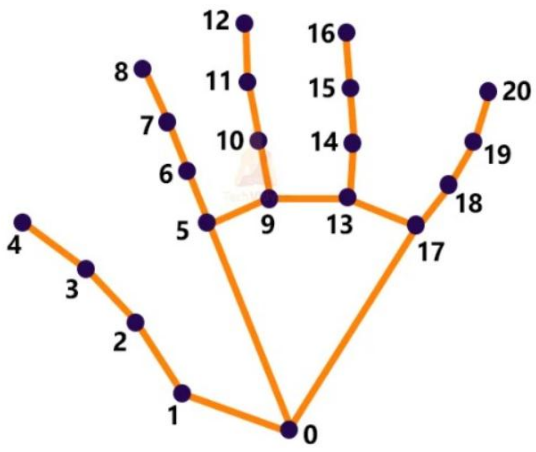


Fig -1: Gesture Key Points

1.4 Holistic Mediapipe Framework

Media Pipe is a custom-designed framework for machine learning solutions developed by Google. It is a frame with an open source and a cross platform, and it is very simple. Media Pipe comes with other pre-trained ML solutions such as face detection, position measurement, hand detection, object detection, etc. It is a reliable tracking of the hand and fingers solution. It uses machine learning (ML) to understand 21 3D local hand marks from just one frame. Although modern methods rely primarily on dynamic desktops to consider, our approach achieves real-time performance on mobile phones, and even scales have gone hand in hand. We hope that providing this visual aid in a comprehensive research and development community will lead to the emergence of creative applications, which encourages new applications and new research methods.

1.5 Objective

The aim of this project was to build a viable Recurrent neural network(RNN) that distinguishes words that are trained in real time by a picture of a signed hand. In this project the initial task is building a language translation System, which can take hand gestures and translate it into written and oral language. Such a translation can greatly reduce the barrier for many people who are deaf and mute so that they can better communicate and others in daily communications among them.

1.6 Summary

Developing a sign language application for the deaf can be very important, as they will be able to communicate easily with those who do not understand sign language. Our program aims to take basic steps to close the gap between ordinary people and the deaf and mute using sign language. The main focus of this work is to develop a vision-based system for identifying spelled ASL characters. The reason for choosing a theory-based system is related to the fact that it provides a simple and accurate way of communicating between a person and a computer. In this report, the various 36:26 sections were classified as English Alphabets. We used Google's Mediapipe Framework Media-pipe Solutions to develop a hand recognition model and can access 3D Three-dimensional

2. LITERATURE SURVEY

In recent years, much research has been conducted on Hand Gesture recognition. The Recognition methods are divided into 2 subparts: -

2.1 Approach Based on Pose Estimation

Real-Time Sign Language Detection using Human Pose Estimation paper, here video conferencing technique is used because it's important at this time. Using methods like Input Representation, Temporal Model, Experiments, and Results, they use the Linear-1 model's coefficients magnitude to visualize how the different landmarks contribute for the pose estimation, they use only returns 17 points compared to the 25 of Open Pose; hence, we map the 17 points to the corresponding indexes for Open Pose. We then normalize the body pose vector by the mean shoulder width the person had in the past 50 frames in order to disregard camera resolution and distance of the signer from the camera.

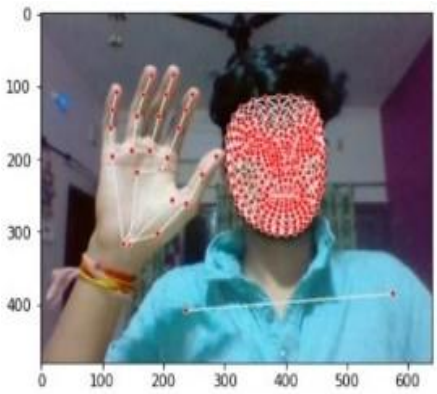


Fig -2: Pose Estimation

2.2 Real Time Recognition Using LSTM Approach

The proposed System in the gesture recognition has been divided into 2 key points: static and dynamic. Static gestures are those in which there is no movement of the hands while performing the signs. Most of the gestures for the alphabets in ISL are static signs. Dynamic gestures, on the other hand, involve hand movements while performing the gestures. RNN model mostly passes the information selectively which may sometimes have an inaccuracy value while computing so here they have used the LSTM model. LSTM is a recurrent network architecture that can learn to bridge time intervals

in excess of steps even in case of noisy, incompressible input sequences, without loss of short time lag capabilities. The nodes are used to read values from the hand gesture. Each of these nodes reads the gesture's accurate value and corresponds to the LSTM which then results in the output.

3. IMPLEMENTATION METHODOLOGY

Many solutions use vision-based systems to track hands, but such a method has many limitations. Users have to move their hands to a limited area, and these systems struggle when hands are crossed or not fully visible. By tracking gesture-based movements, however, gesture recognition systems are able to detect both vertical and shifting touches in real-time. The system separates the hand and the back using color and depth of the data. The gesture collected as a sample is then further divided into a palm, wrist, and fingers in the form of nodes. The system then ignores the arm and wrist as it does not provide meaningful information.

Hence, the system deceives information about the distance of the image within the fingers to the center of the hand, finger height, finger position, and more. At the end, the system recalls all the extracts from the gesture representing the hand image.

The flow diagram for the Real-time sign Recognition using LSTM is depicted in Fig. 3,

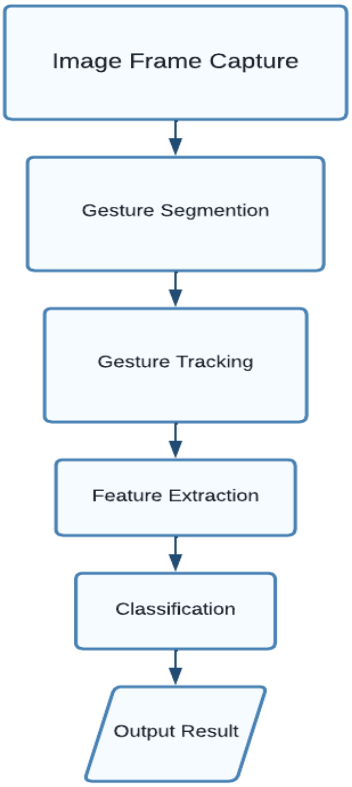
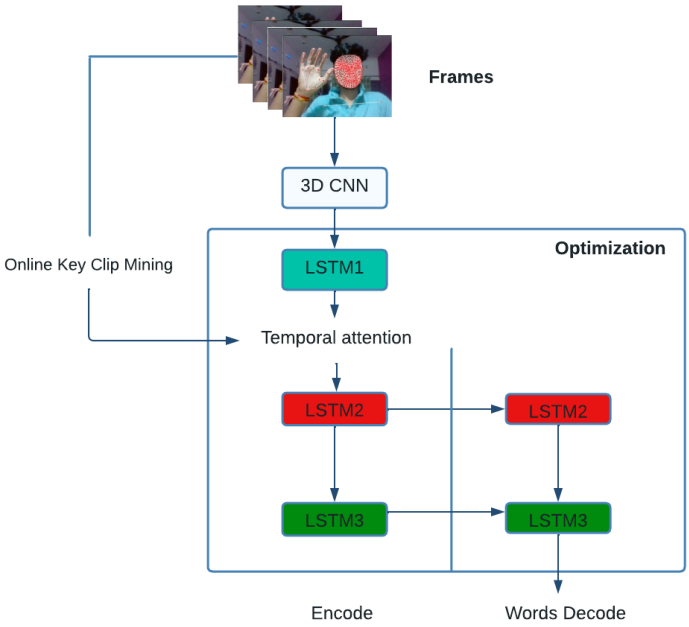


Fig -3: Flow Chart

3.1 Gesture Recognition Using LSTM

The Network that has been proposed includes a 3-Dimensional convolution neural network also called 3DCNN followed by the Long-short-term-memory (LSTM) as a feature-release model. The Multimodal data along with RGB and depth data are integrated as an input model. Also, the Finite State Machine (FSM) control is used to limit the flow of several key points and to limit the alert phases. The use of small classes is often shown with higher accuracy than that of the large classes. The re-use of multiple commands in a single application totally depends on the context of each application itself. Within the system, the global side of hand touch detection is explored. Blocked by the scope of the action, instead of using the finger element to distinguish, the whole handshape as data entry is done. Attention is focused on the hand by removing the background and unnecessary pixels. This method allows the system to catch in any variety of situations, e.g., overcrowded areas, and may speed up model calculation.



4. Conclusion

In this project, we have implemented an automatic sign language recognition system in real-time, using tools learned in computer vision and machine learning. We learned about how sometimes basic approaches work better than complicated approaches. Despite trying to use a smart segmentation algorithm, the relatively basic skin segmentation model turned out to extract the best skin masks. We also realized the time constraints and difficulties of creating a dataset from the scratch.

5. Hardware Requirements

Windows computer or Linux, Python installed along with the libraries.

1) System Webcam
2) Python Version 3.9
3) Intel Core i5 9th Gen.
4) GPU used - Nvidia GTX 1050 Ti
5) 720p60 Web Camera
6) Python 3.8.6 and IDE like Jupiter Notebook, PyCharm etc.
7) Libraries: OpenCV, TensorFlow, Matplotlib, Media-pipe, and many more basic *RNN from the Scikit Library of Python

6. References

[1] A. Mittal, P. Kumar, P. P. Roy, R. Balasubramanian, and B. B. Chaudhuri, "A Modified LSTM Model for Continuous Sign Language Recognition Using Leap Motion," in IEEE Sensors Journal, vol. 19, no. 16, pp. 7056-7063, 15 Aug.15, 2019.

[2] P. Likhar, N. K. Bhagat and R. G N, "Deep Learning Methods for Indian Sign Language Recognition," 2020 IEEE 10th International Conference on Consumer Electronics (ICCE-Berlin), 2020, pp. 1-6.

[3] A. Chaikaew, K. Somkuan and T. Yuyen, "Thai Sign Language Recognition: an Application of Deep Neural Network," 2021 Joint International Conference on Digital Arts, Media and Technology with ECTI Northern Section Conference on Electrical, Electronics, Computer and Telecommunication Engineering, 2021, pp. 128-131.

[4 A. Mittal, P. Kumar, P. P. Roy, R. Balasubramanian and B. B. Chaudhuri, "A Modified LSTM Model for Continuous Sign Language Recognition Using Leap Motion," in IEEE Sensors Journal, vol. 19, no. 16, pp. 7056-7063, 15 Aug.15, 2019.

[5] S.Aly and W. Aly, "DeepArSLR: A Novel Signer-Independent Deep Learning Framework for Isolated Arabic Sign Language Gestures Recognition," in IEEE Access, vol. 8, pp. 83199-83212, 2020.

[6] Cheok, M.J., Omar, Z. & Jaward, M.H. A review of hand gesture and sign language recognition techniques. Int. J. Mach. Learn. & Cyber. 10, 131–153.

[7] Moryossef, A., Tsochantaridis, I., Aharoni, R., Ebling, S., Narayanan, S. (2020). Real-Time Sign Language Detection Using Human Pose Estimation. In: Bartoli, A., Fusiello, A. (eds) Computer Vision – ECCV 2020 Workshops. ECCV 2020. Lecture Notes in Computer Science, vol 12536. Springer.

[8] Mehreen Hurroo , Mohammad Elham, 2020, Sign Language Recognition System using Convolutional Neural Network and Computer Vision, INTERNATIONAL JOURNAL OF ENGINEERING RESEARCH & TECHNOLOGY (IJERT) Volume 09, Issue 12 (December 2020).

[9] Muthu Mariappan, H and V DrGomathi. "Indian Sign Language Recognition through Hybrid ConvNet-LSTM Networks." EMITTER International Journal of Engineering Technology (2021).

[10] Shivashankara, Ss, and S. Srinath. "American sign language recognition system: an optimal approach." International Journal of Image, Graphics and Signal Processing 11, no. 8 (2018): 18.

[11] Science, WARSE The World Academy of Research in, and Engineering. "Sign Language and Common Gesture Using CNN." International Journal of Advanced Trends in Computer Science and Engineering, 2021.

[12] Mrs.Dipali Rojasara, Dr.Nehal G Chitaliya, 2013, Indian Sign Language Recognition – A Survey, INTERNATIONAL JOURNAL OF ENGINEERING RESEARCH & TECHNOLOGY (IJERT) Volume 02, Issue 10 (October 2013),

[13] Chava Sri Varshini , Gurram Hruday , Mysakshi Chandu, Guguloth Sreekanth , Shaik Khadar Sharif, 2020, Sign Language Recognition, INTERNATIONAL JOURNAL OF ENGINEERING RESEARCH & TECHNOLOGY (IJERT) Volume 09, Issue 05 (May 2020),

[14] D. Avola, M. Bernardi, L. Cinque, G. L. Foresti and C. Massaroni, "Exploiting Recurrent Neural Networks and Leap Motion Controller for the Recognition of Sign Language and Semaphoric Hand Gestures," in IEEE Transactions on Multimedia, vol. 21, no. 1, pp. 234-245, Jan. 2019,