Available online at http://www.advancedscientificjournal.com
http://www.krishmapublication.com
IJMASRI, Vol. 2, issue 6, pp. 676- 681, July -2022
10.53633/ijmasri.2022.2.6.001

INTERNATIONAL JOURNAL OF MULTIDISCIPLINARY ADVANCED SCIENTIFIC RESEARCH AND INNOVATION (IJMASRI)

ISSN: 2582-9130 **IBI IMPACT FACTOR 1.5** **DOI: 10.53633/IJMASRI**

RESEARCH ARTICLE

SIGN LANGUAGE RECOGNITION IN REAL-TIME

Yash Saxena
Department of Computer Science Engineering, Maharaja Agrasen Institute of Technology, Rohini, Delhi
Email : yashsaxena.mait@gmail.com

Abstract

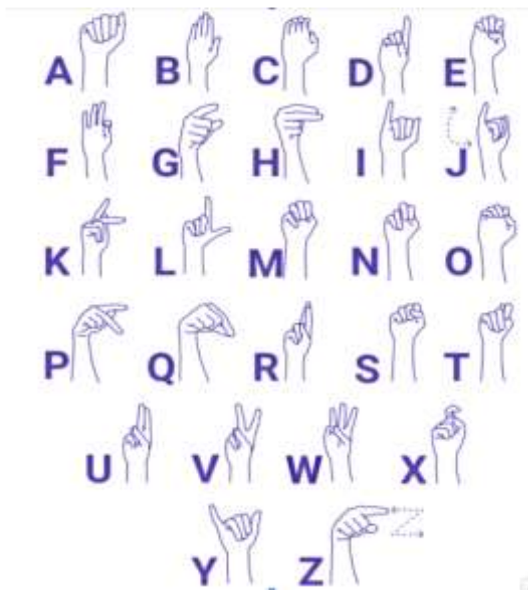
Every day we cross many people who are facing a hard life because of being deaf and dumb. There are not many technologies to help them communicate and interact with everyone. They face hardships that are not easy to overcome even in today's modern era. Hence, to address a portion of their shortcomings and enable a way for them to communicate with everyone else now it's time to opt for a real-time hand tracking on-device solution. For communicating only with gestures and not voice, hand-pose-based Sign Language will be a great option, and hence detecting hand poses and gestures can be carried out by many innovative methods like RNN, CNN, sensor gloves, SVM, etc. The proposed technique suggested in this paper uses a landmark model that can predict the hand skeleton by detecting 21 hand landmarks concerning the structure of a hand. On the basis of the literature survey, the results of hand pose detection were quite accurate, and it is a big step toward our goal.

Keywords: Image processing, ASL recognition system, Convolution neural network (CNNs), Classification, real time

Introduction

Hand recognition and tracking is a crucial component for achieving a natural way of interaction and communication solely based on gestures and not voice. This paper's purpose is to review various different techniques to determine which letter of the "American Sign Language (ASL)" alphabet is being signed using a picture or a real-time video of a signing hand. This study is a collection of various steps toward the development of a hypothetical sign language translator that can transform hand gesture-based linguistic signals into written and oral language. This type of translator would drastically reduce the communication barrier for mute and deaf people, allowing them to converse more effectively in everyday situations.

Sign Language is accompanied by fingerspelling. Fingerspelling, which can spell out words character by character, and word level association, which incorporates hand movements that transmit word meaning, are two types of sign language. Fingerspelling is an important technique in sign language because it allows people to communicate names, addresses, and other words that don't have significance at the word level. Despite this, fingerspelling is not frequently utilized since it is difficult to grasp and use. Furthermore, there is no universal sign language.



Literature Survey

- Sahin and Savur, 2015) proposed an armband recognition system for American Sign Language. The results obtained in real-time were 82.3% accurate and used 2080 samples in total for sign alphabets. The classification of samples was done using SVM.

- (Sun C, Bao BK, Xu, Zhang T 2013) have worked and developed Vector Machine with Latent Support to classify sentences. They extracted Histogram of Oriented Gradient(HOG) as well as Kinect and other optic flow features for sign recognition and were able to achieve an accuracy of 86% for continuous American signs.

- (Oz Cand Leu MC, 2013)developed a sensory gloves-based solution for hand pose recognition and were able to convert poses to ASL English words. they used static samples with a single hand from 50 words and the neural network classification method and were able to obtain an accuracy of 90%.

- (Zamani and Kanan 2014) build a recognition solution based on still pictures to determine American alphabets and symbols. A dataset of 2520 picture samples was collected of single-hand signs and used Neural Network for recognition. The model supported only static gestures and was able to reach an accuracy of 99.88%.

- (Lionel P, Sander D, Pieter-Jan K, Benjamin S .2015) carried out extensive research to develop a sign language recognition solution using Recurrent Neural Network and Convolution Neural Network. The author used a dataset of 20 distinct Italian gestures signed in different surroundings by an audience of 27 people and was able to automate the recognition process and divide the process into two steps - feature extraction and action classification. RNN, CNN, and Artificial Intelligence were combined to perform the steps.

- (Wu J, Sun L, Jafar, Tian, Estevez 2015) suggested an electromyography and arm sensors-based recognition system for sign language. They collected a dataset of 40 signs and used four classifiers to classify them. Classification methods were LibSVM, Naïve Bayes, Decision tree, and Nearest-Neighbor. Many comparative results were deduced and SVM came out with the highest degree of accuracy.

- (Naglot and Kulkarni 2016) proposed a solution based on ANN using Leap Motion Controller for Indian Sign Language recognition. Multilayer Perceptron (MLP) was used as a classification method for dynamic single-handed gestures. An accuracy of 100% was achieved.

- (Sagar P. More, Abdul Sattar, 2015) presented a system for static hand gesture recognition. The classification method used was digital image processing and SIFT algorithm for gesture features. SIFT features are calculated at fixed edges, rotation, audio addition.

TABLE 1

Summary of Literature Survey and comparisons

Author	**Acquisition mode**	**Signing mode**	**Static/ Dynamic gesture**	**Single/ Double handed**	**Technique used**	**Recognition rate**
Sahin, Savur	Armband	Isolated sign	Both	Single	SVM	82.3%
Sun C, Bao BK, Xu, Zhang T	Kinect	Continuous signs	Dynamic	Double	Latend support vector-Machine	86%
Oz C, Leu MC	Glove based	Isolated sign	Static	Single	Neural Network	90%
Sagar P. More, Abdul Sattar	Camera	Isolated sign	Static	Single	Image Processing and SIFT	>90%
Nakul Nagpal, Dr. Arun Mitra, Dr. Pankaj Agrawal	Camera	Isolated sign	Both	Double	Neural Network	>90%
Zamani, Kanan	Camera	Isolated sign	Static	Single	Neural Network	99.88%
Lionel P, Sander D, Pieter-Jan K, Benjamin S	Camera	Isolated sign	Static	Single	CNN and RNN	93.75%
Kulkarni, Naglot	Leap motion	Isolated sign	Dynamic	Single	ANN	100%
Wu J, Sun L, Jafar, Tian, Estevez	Arm Sensors	Isolated sign	Dynamic	Single	SVM, Naïve Bayes, Decision Tree, Nearest-Neighbor	99.09% 84.11% 81.88% 98.96%

Proposed architecture

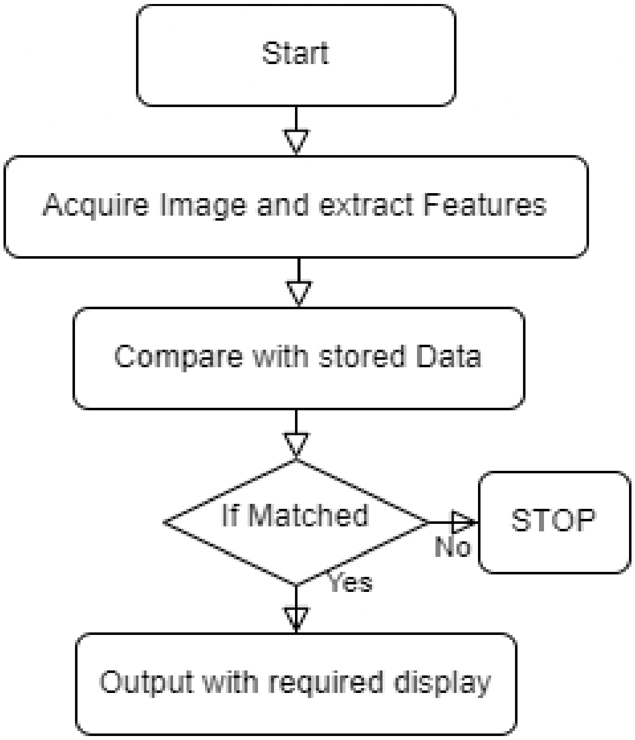


Fig. 1: Proposed Architecture Flowchart

A. Palm Detection

As per the task of detecting hand pose to further process the structure and recognize hand gestures, we first need a palm detection model. We can first detect the initial location of the hand by employing a model similar to Blazeface (2019) created for face detection. Since hand sizes vary by a wide variety and detecting them needs model training with keeping that in check. We can estimate the square bounding box over the palm and reduce the number of anchors by ignoring other aspect ratios

Fig. 2: Palm Detection (BlazePalm detection)

B. Hand Landmark Detection

We first run a palm detection, after running the model on the complete image. The hand landmark model should enable accurate landmark localization of a total of 21 2.5D coordinates within the regression-detected hand region(s). The model should learn steady internal hand-pose description and should be able to even detect partially visible hands and folded orientations of the hand (2020).

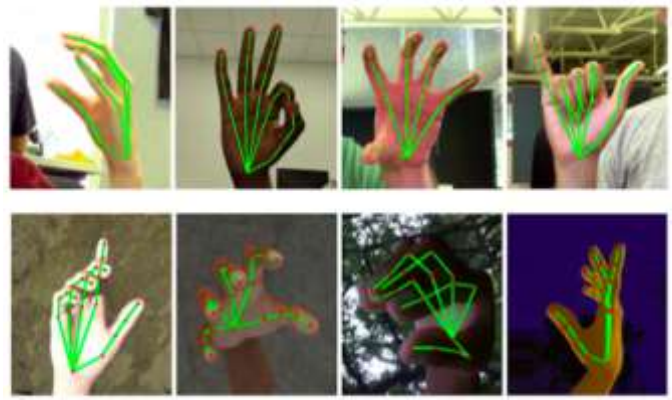


Fig. 3: On top detection of real hands and below detection of synthetic hands. Google[19]

The model will have three outputs:-

- To detect 21 unique landmarks for the detection of the hand's structure based on the action skeletal structure of the human hand. The landmarks should have x and y coordinates and one depth coordinate relative to the picture based on depth estimation.
- The probability output for the presence of hands in the input feed/image.
- A binary flag to determine hand, i.e., left or right hand.

The 21 landmarks proposed should use the same topology as (Tomas Simon. 2017). There a be tracking failures and to recover from those failures, a probability output mechanism is also essential for the model to determine whether or not an aligned hand structure exists in the given area, and in case of lower score than a fixed probability threshold, the detection mechanism can be re-triggered to calculate the tracking again. Handedness is also a very crucial requirement for hand pose-related interactions. This is particularly handy in a few situations where each hand has its own set of functions. As a result, the model has a binary classification flag to forecast the input hand's type, i.e., left or right.

Gesture Recognition

We use the 21 2.5D (3D) coordinates which are generated by the model as inputs for gesture recognition. Each corresponds to one point on the hand. Thus for gesture recognition, we get 63(21 x 3)D input vector representation. To define the gesture, these coordinate vectors are then translated into the features that the neural network expects. There are two kinds of gestures, static gestures and dynamic gestures. The current architecture is proposed by taking static gestures as the focus of concern.

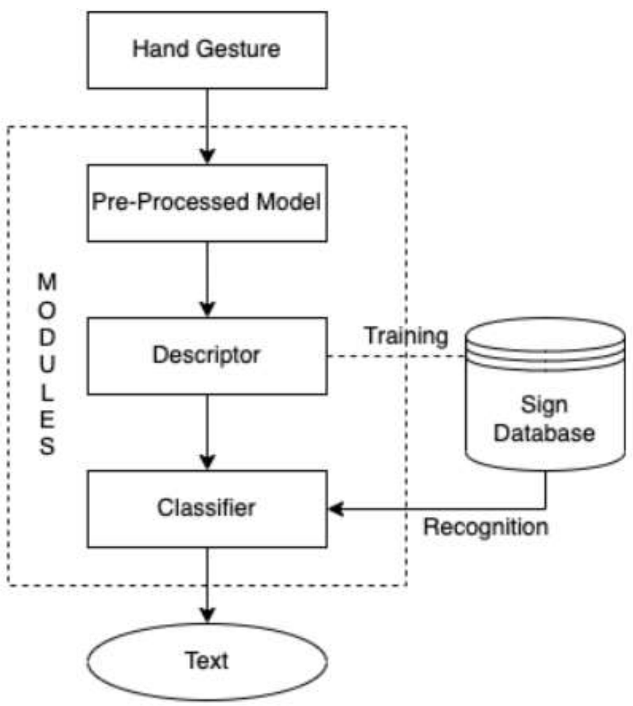


Fig. 4: Static Gesture Recognition

Static Gestures: Static gestures are ones in which only one hand-pose is used to depict the gesture. The current model can recognize handedness and which hand is being utilized, allowing us to determine static gestures based on this information.

The gestures can be determined by comparing the coordinate vectors with estimated vectors to compare the similarities in hand pose and then check if transforming the vectors can match with the predefined gesture.

Application Examples

This Hand tracking model for gesture recognition can readily be used in many day-to-day life scenarios as well as for professional purposes. After predicting the skeleton structure of the hand, we can compare gestures using some simple algorithms to compute similarities. The angles of joints accumulated, for example, can indicate the status of each finger, such as whether it is bent or straight. We may also map a collection of predefined gestures to a set of finger(s) states. This simple but powerful strategy allows us to accurately predict even the most basic static gestures. Along with static gestures recognition, it is also possible to utilize the landmark sequence to estimate various dynamic gestures.

Conclusion

In this paper, we have reviewed and discussed various end-to-end hand tracking solutions and research works as well as we've seen a suitable way to recognize and determine hand gestures using 21 landmarks. We have discussed how hands can be divided and represented as a mesh of 21 2.5D points which can be considered the main landmarks for the human hand structure. We can also differentiate among hands using the handiness binary classification. The main goal of this paper was to look at how hand poses can be determined using the CNN provided by the TensorFlow hand pose model and utilizing it for sign language detection.

Reference

1. Savur and Sahin ((2015) "Real-time American sign language recognition system using surface EMG signal". ieee 14th icmla, pp 497–450.
2. Sun, C., Zhang, T. Bao, B.K. Xu, C.(2013). "Latent support vector machine for sign

3. language recognition with Kinect". IEEE 20th ICIP, pp 4190–4194.
4. Oz, C and Leu, M.C. (2011), "American Sign Language word recognition with a sensory glove using artificial neural networks". Eng Appl Artif Intell 24(7): 1204–1213.
5. Sagar P.More and Abdul Sattar, (2015). "Hand gesture recognition system for dumb people". IJEET, vol.3, Issue2.
6. Nakul Nagpal., Arun Mitra and Pankaj Agrawal(2021). "Design Issue and Proposed Implementation of Communication Aid for Deaf Dumb People", International Journal on Recent and Innovation Trends in Computing and Communication, Volume: 3 Issue: 5, pp-147 – 149.
7. Zamani M and Kanan H.R. (2014). "Saliency based alphabet and numbers of American Sign Language recognition using linear feature extraction".IEEE 4th International E-Conference on Computer and Knowledge Engineering(ICCKE), pp 398–403.
8. Lionel, P., Sander, D. Pieter-Jan, K and Benjamin S. (2015) "Sign Language Recognition using Convolutional Neural Networks", presented at the Ghent University, ELIS, Belgium. DOI:10.1007/978-3-319-16178-5 40.
9. Wu, J., Tian, Z. Sun, L. Estevez, L and Jafari, R,.(2015)."Real-time American Sign Language recognition using wrist-worn motion and surface EMG sensors". IEEE 12th International Conference on Wearable and Implantable Body Sensor Networks(BSN), pp 1–6.
10. Kulkarni, M and Naglot, D. (2016). "ANN-based Indian Sign Language numerals recognition using the leap motion controller". IEEE 2016 ICICT, vol 2, pp 1–6.
11. Valentin, B., Yury, K. Andrey, V. Karthik, R and Matthias, G.(2019). "Blazeface: Sub-millisecond neural face detection on mobile GPUs". CoRR, abs/1907.05047.
12. Fan Zhang., Valentin Bazarevsky, Andrey Vakunov, Andrei Tkachenka, George Sung, Chuo-Ling Chang and Matthias Grundmann. (2020) .Google Research. "MediaPipe Hands: On-device Real-time Hand Tracking" arXiv:2006.10214v, 2020.
13. Tomas Simon., Hanbyul Joo, Iain, A. Matthews, and Yaser Sheikh. (2017). Hand keypoint detection in single images using multiview bootstrapping. CoRR, abs/1704.07809.
14. Pratibha Pandey and Vinay Jain, (2015). "Hand Gesture Recognition for Sign Language Recognition: A Review", International Journal of Science, Engineering and Technology Research (IJSETR), Volume 4, Issue 3, March 2015.
15. Zafar Ahmed Ansari and Gaurav Harit, (2016). "Nearest Neighbour Classification of Indian Sign Language Gestures using Kinect Camera", in Sadhana, Vol. 41, No. 2, February 2016, pp. 161-182.
16. Divya Deora and Nikesh Bajaj (2012). "Indian Sign Language Recognition", in Emerging Technology Trends in Electronics, Communication, and Networking (ET2ECN), 2012 1st International Conf.DOI: 10.1109/ET2ECN.2012.6470093
17. Chenyang Zhang (2016) "Multi-modality American Sign Language Recognition", in Image Processing (ICIP), 2016 IEEE International Conf. DOI: 10.1109/ICIP.2016.7532886.
18. Tung-Lin "Dove" Wu and Taishi "Wabi" Senda.(2021). "Pen Spinning Hand Movement Analysis Using MediaPipe Hands" arXiv:2108.10716v1, 24 Aug 2021.
19. Sreehari (2016) "Weekend project: sign language and static gesture recognition using scikit-learn". Available: https://medium.freecodecamp.org/weekend-projects-sign-languageandstatic-gesture-recognition-using-scikit-learn-60813d600e79.
20. Google Home - Mediapipe. 2020. URL: https://google.github.io/mediapipe/

*****