International Research Journal of Engineering and Technology (IRJET) e-ISSN: 2395-0056
Volume: 07 Issue: 12 | Dec 2020 www.irjet.net p-ISSN: 2395-0072

Sign Language Recognition Techniques: A survey

Omkar Govalkar[1], Pratik Gaikar[2], Pramod Gavali[3]

[1,2,3]*Dept. of Computer Engineering, Smt. Indira Gandhi College of Engineering, Navi Mumbai, Maharashtra, India*

Abstract - *From the past few years, Sign language recognition has emerged as one of the important areas of research in computer vision. As the community of speech and hearing-impaired people has depended on sign language as a communication medium. New techniques have been developed from past decades till now, to counter the problem of building a communication bridge between normal people and speech and hearing-impaired people. Sign gestures can be classified as static and dynamic. However static gesture recognition is simpler than dynamic gesture recognition but both recognition systems are important to the human community. Efficient sign language recognition techniques were studied and summarized in this paper.*

***Key Words*: Sign language recognition, Indian Sign Language, Vision-based, Neural Network(NN), Human Computer Interaction(HCI), Region of Interest(ROI)**

1. INTRODUCTION

Speech and hearing impaired people are humans at the deepest psychological level. Many of these people are not even exposed to sign languages and it is observed that it gives a great relief on a psychological level, when they find out about signing to connect themselves with others by expressing their love or emotions. About 5% population in world are suffering from hearing loss. Speech and hearing impaired people use sign language as their primary means to express their thoughts and ideas to the people around them with different hand and body gestures. There are only about 250 certified sign language interpreters in India for a deaf population of around 7 million. In this work, the design of prototype of an assistive device for speech and hearing-impaired people is presented so as to reduce this communication gap with the normal people. This device is portable and can hang over the neck. This device allows the person to communicate with sign hand postures in order to recognize different gestures based signs. The controller of this assistive device is developed for processing the images of gestures by employing various image processing techniques and deep learning models to recognize the sign. This sign is converted into speech in real-time using text-to-speech module[4].

Hand gesture recognition provides an intelligent, natural, and convenient way of human–computer interaction (HCI). Sign language recognition (SLR) and gesture-based control are two major applications for hand gesture recognition technologies. SLR aims to interpret sign languages automatically by a computer in order to help the speech and hearing-impaired communicate with hearing society conveniently. Since sign language is a kind of highly structured and largely symbolic human gesture set, SLR also serves as a good basic for the development of general gesture-based HCI. Speech and hearing-impaired use hand signs to communicate, hence normal people face problem in recognizing their language by signs made. Hence there is a need of the systems which recognizes the different signs and conveys the information to the normal people.

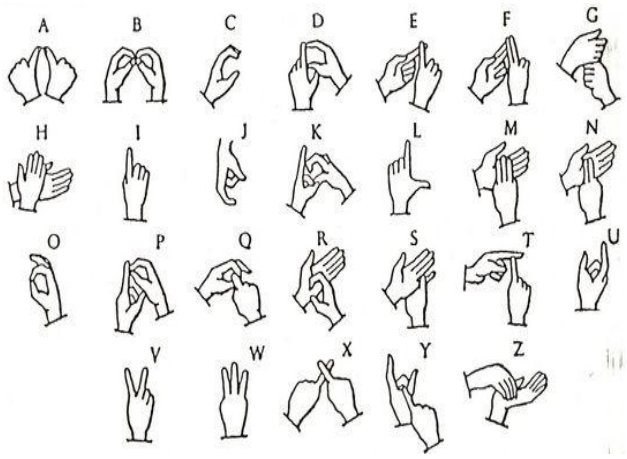

Fig.1**:** Indian Sign Language representing letters of the English alphabet.

2. LITERATURE SURVEY:

2.1. SIGN LANGUAGE DATASETS:

Sign language consists of much more than just the hand signs; it also includes varieties of features such as hands, facial expressions, mouth and body position. Sign language divides in two parts i.e static and dynamic gestures. Dynamic signs exhibit motion, whereas static signs are characterized by a specific static posture[3]. Fig.1 shows the Indian sign language hand posture for all alphabets. It is difficult to extract features from the available dataset, we can create our own raw dataset according to our needs.

2.2. SIGN LANGUAGE RECOGNITION:

There appears to be several possible method for recognition of a hand sign through different approaches.

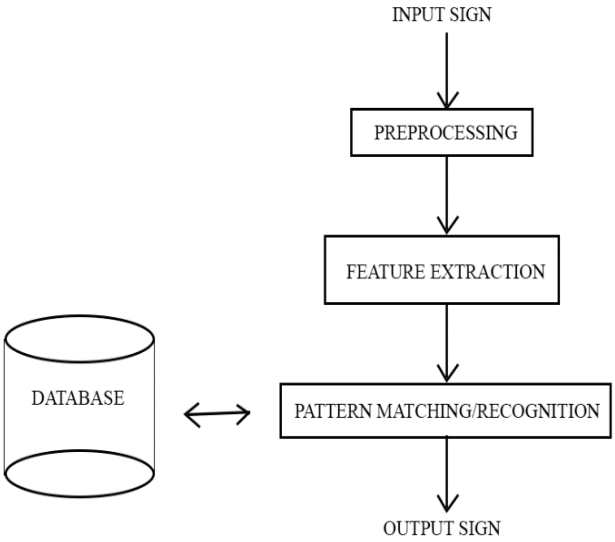

Fig.2: Basic System Architecture

The basic flow structure of any sign language reognition would looks like this 3 steps process. First, the input data is acquired, for example via the phone camera or from some sensor. The next step is to extract the features from the input data. Finally, the sign is classified using some appropriate algorithm that is compatible with the extracted features. For each method we examine, we take a close look in how that method approaches the problems of feature extraction and recognition/classification.

2.2.1 Sensor Based Approach:

This approach collects the data of gesture performed by using different sensors. The data is then analyzed and conclusions are drawn in accordance with the recognition model. In case of hand gesture recognition different types of sensors were used and placed on hand, when the hand performs any gesture, the data is recorded and is then further analyzed. The first sensor used was Data gloves then LED's came into existence. The invention of the first data glove was done in 1977. Sensor based approach damages the natural motion of hand because of use of external hardware. The major disadvantage is complex gestures cannot be performed using this method[3].

2.2.1.1 Using Gloves:

Glove-based approaches have been implemented using sensors that track hand gestures. Multiple sensors embedded in the gloves are used to track the fingers, palm and their location and motion. Such an approach provides coordinates of the palm and fingers for further processing. These devices may be connected wirelessly via Bluetooth[3].

2.2.1.2 Microsoft Kinect:

Microsoft Kinect tracks your full body movement in 3-D, while responding to commands, directions, even a difference of emotion in your voice. The 3D depth information from hand motions, generated from Microsoft's Kinect sensor and apply a hierarchical conditional random field (CRF) that recognizes hand signs from the hand motions. The proposed method uses a hierarchical CRF to detect candidate segments of signs using hand motions, and then a BoostMap embedding method to verify the hand shapes of the segmented signs[8].

2.2.1.3 Smartphone Internal Sensors:

Smartphone internal sensors Approach Recently, new smartphones have been embedded with sensors that help to detect the posture and motion of the device. Numerous researchers utilize this feature to create gesture recognition models. The main issue with this approach is the limitation of signs details provided by the sensors[3].

2.2.2 Vision Based Approach:

This approach takes image from the camera as data of gesture. The vision based method mainly concentrates on captured image of gesture and extract the main feature and recognizes it[3]. It concentrates on region of interest from the realtime video and extracts only that part and then various filters are used to correctly classify the structure of a hand sign. Skin segmentation algorithms, which often depend on specifying thresholds [18], are widely used in Computer Vision applications. Extracting accurate hand features is a major challenge for the vision-based approach. Extraction is affected by many factors, such as lighting condition and background noise. The more accurate the detection and extraction is, the better the recognition results become. Some systems assume that the only visible object in the captured image is the hand [5, 2], which uses cascades of boosted rectangle filters, is a well-known method, that is commonly used for detecting hands. Some researchers [5, 2] implement the Viola-Jones method on portable platforms, as Viola-Jones is relatively easy to implement and has low hardware requirements. Due to slow processing time in some models, a clientserver framework is used. In such a framework, the phone is connected to a regular computer via wireless network.

The main goal of vision based system is to capture image or gesture and that can be done in two ways i.e. static and dynamic gesture.

2.2.2.1 Static gesture:

Static gesture recognizing aims at identify particular kinds of posture which remain still for a period of time in the videos. Static gestures are single images which involves no time frame. We use adaboost algorithm when training cascade classifiers, which can promote systems to be robust and real-time and can be used to identify static gestures[6].

2.2.2.1.1 DenseNet :

DenseNet has been widely used for classification tasks due to the advantages it introduces such as alleviating the vanishing gradient – a common problem encountered with deep networks.

DenseNet has been widely used for classification tasks due to the advantages it introduces such as alleviating the vanishing gradient – a common problem encountered with deep networks.

2.2.2.1.2 Support Vector Machine (SVM) :

A support vector machine (SVM) is a supervised machine learning model that uses classification algorithms for two-group classification problems. After giving an SVM model sets of labeled training data for each category, they're able to categorize new given problem.

2.2.2.2 Dynamic gesture :

Dynamic gesture recognizing is mainly focusing on recording the position of hands in images and identifies the meaning of actions by analyzing them. The first step is to identify hand automatically, that is to say, the system itself can get the position of the hand in images without manual intervention. We can utilize the method used in static gesture recognizing to judge if any hands is in images. In sign laguage recognition system, gestures cannot be judged at the time they are detected, since the dynamic gestures should be taken into consideration. Namely, we do not judge gesture only by one image but just record the position of hand because it may be one fragment in a series of images which contents a dynamic gesture such as a wave[6]. Efficient dynamic gesture recognition techniques are studied in the following:

2.2.2.2.1 Hidden Markov Model (HMM):

After the trajectory is obtained from the tracking algorithm, features are abstracted and used to compute the probability of each gesture type with Hidden Markov Model. We use a vector to describe those features and as the input of the Hidden Markov Model. Hidden Markov Model is a statistical Markov model in which the system being modeled is assumed to be a Markov process – call it – with unobservable ("hidden") states. Hidden Markov Model assumes that there is another process whose behavior "depends" on. The goal is to learn about by observing.

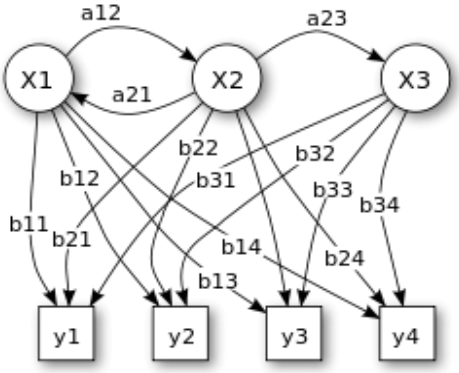

Fig.3 : Probabilistic parameters of a hidden Markov model

2.2.2.2.2 Openpose Estimation:

Realtime multi-person 2D pose estimation is a key component in enabling machines to have an understanding of people in images and videos. We can present a realtime approach to detect the 2D pose of multiple people in an image. The proposed method uses a nonparametric representation, which we refer to as Part Affinity Fields (PAFs), to learn to associate body parts with individuals in the image. This bottom-up system achieves high accuracy and realtime performance, regardless of the number of people in the image[1].

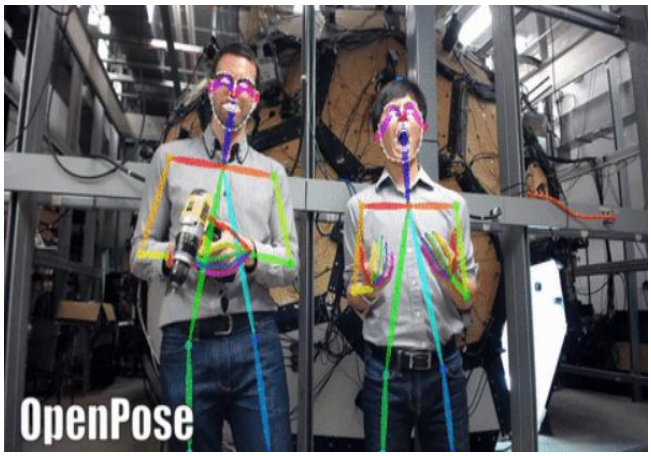

Fig.4 : Real-time multi-person system to jointly detect human body, hand, facial, and foot keypoints (in total 135 keypoints) on single images[1].

2.2.2.2.3 Mediapipe :

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 method achieves real-time performance on a mobile phone, and even scales to multiple hands. Mediapipe hands utilizes an ML pipeline consisting of multiple models working together: A palm detection model that operates on the full image and returns an oriented hand bounding box. A hand landmark model that operates on the cropped image region defined by the palm detector and returns high-fidelity 3D hand keypoints[7].

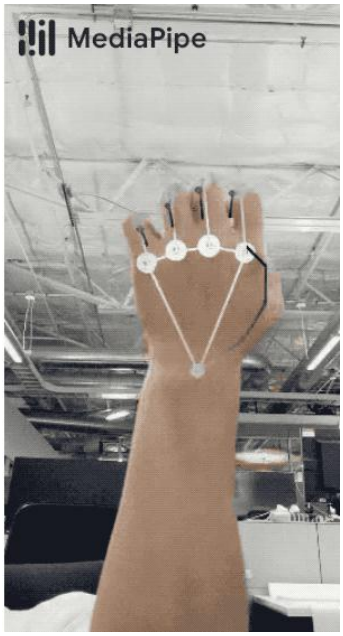

Fig.5 : Tracked 3D hand landmarks are represented by dots in different shades, with the brighter ones denoting landmarks closer to the camera[7].

3. CONCLUSIONS

In this paper we have listed out several techniques for the real-time sign language recognition for both sign nature i.e. static gesture and dynamic gesture with the few techniques in both. The accuracy and the efficiency can be improved using one of this methods for realtime sign language recognition.

4. REFERENCES

[1] Z. Cao, G. Hidalgo, T. Simon, S. -E. Wei and Y. Sheikh, "OpenPose: Realtime Multi-Person 2D Pose Estimation Using Part Affinity Fields," in IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 43, no. 1, pp. 172-186, 1 Jan. 2021, doi: 10.1109/TPAMI.2019.2929257.

[2]H. Elleuch, A. Wali, A. Samet and A. M. Alimi, "A static hand gesture recognition system for real time mobile device monitoring," 2015 15th International Conference on Intelligent Systems Design and Applications (ISDA), Marrakech, 2015, pp. 195-200, doi: 10.1109/ISDA.2015.7489224.

[3] Ghanem, Sakher & Conly, Christopher & Athitsos, Vassilis. (2017). A Survey on Sign Language Recognition Using Smartphones. 171-176. 10.1145/3056540.3056549.

[4] L. Boppana, R. Ahamed, H. Rane and R. K. Kodali, "Assistive Sign Language Converter for Deaf and Dumb," 2019 International Conference on Internet of Things (iThings) and IEEE Green Computing and Communications (GreenCom) and IEEE Cyber, Physical and Social Computing (CPSCom) and IEEE Smart Data (SmartData), Atlanta, GA, USA, 2019, pp. 302-307, doi: 10.1109/iThings/GreenCom/CPSCom/SmartData.2019.00071.

[5] R. Y. Hakkun, A. Baharuddin, et al. Sign language learning based on android for deaf and speech impaired people. In Electronics Symposium (IES), 2015 International, pages 114–117. IEEE, 2015.

[6] Zong-yuan Zhao, Wan-lin Gao, Miao-miao Zhu, Lina Yu, A Vision Based Method to Distinguish and Recognize Static and Dynamic Gesture, Procedia Engineering, Volume 29, 2012,

[7]https://google.github.io/mediapipe/solutions/hands.html

[8] Yang, H.-D. Sign Language Recognition with the Kinect Sensor Based on Conditional Random Fields. *Sensors* 2015, *15*, 135-147.