2020 IEEE International Conference for Innovation in Technology (INOCON)
Bengaluru, India. Nov 6-8, 2020

Sign Language Recognition Techniques- A Review

Mohammed Safeel[1], Tejas Sukumar[2], Shashank K S[3], Arman M D[4], Shashidhar R[5], Puneeth S B[6]
[1,2,3,4] Department of Electronics and Communication, JSS Science and Technology University Mysuru, Mysuru, India
[5]Assistant Professor, Department of Electronics and Communication, JSS Science and Technology University Mysuru, Mysuru, India
[6]Assistant Professor, Department of Electronics and Communication, Presidency University, Bengaluru, Bengaluru, India
[5]shashidhar.r@sjce.ac.in, [6]puneeth003@gmail.com

***Abstract*— Sign language reduces the barrier for communicating with the humans having impaired of speech and hearing, on the other hand Sign language cannot be easily understood by common people. Therefore, a platform is necessary that is built using an algorithm to recognize various signs it is called as Sign Language Recognition (SLR). It is a technique that simplifies the communication between speech and hearing impaired people with normal people, the main aim of SLR is to overcome the aforementioned drawback. In this manuscript it is aimed to review various techniques that have been employed in the recent past for SLR that are employed at various stages of recognition. Adding to the above, various image based with or without the glove employed for detection, their advantages and difficulties encountered during this process. Also, segmentation, feature extraction, methods used for feature vector quantization and reduction techniques are discussed in detail. Along with these, during classification it involves training, testing that employs various training models, including Hidden Markov Model based approaches and Deep learning methods such as CNN, also techniques like k-NN, ANN, SVM and others. Then finally discussion on results and observations from several techniques are compared. The approaches that are being reviewed are so flexible that they can be employed for major sign detections with applications in various domains.**

Keywords*— *Sign language recognition, HMM, Segmentation, Feature extraction, Neural networks, SVM, Recognition.

I. Introduction

Sign language is a natural language that uses vision gestures and signs to convey meaningful information among people. Specially, speech and hearing impaired people use this to communicate among themselves and others by using hand using the combination of gestures, movements and signs. Sign language uses manual as well as non-manual signals. They involve movement of hand, finger, arms, face and neck [1]–[3].

As the whole world do not have a common spoken language similarly there is no universal sign language to be used by everyone, but it is characterized by regions as well as countries. Some of the major sign languages that are most popular are American Sign Language (ASL), French Sign Language (FSL), Indian Sign Language (ISL), and Japanese Sign Language (JSL)[21,24]. Although some of the symbols are quite similar between these sign languages. This helps people who have problems related to voice and are unable to communicate normally with others. Hence, this has generated a need to design a system that allows hearing impaired people to communicate easily with normal people, since most normal people do not have the knowledge of sign language systems. Sign language involves manual as well as non-manual signals (NMS), NMS represent grammatical as well as semantic features other than hands. They usually include eye gazes, face expressions, body shifts and head tilts. However, manual signals (MS) have an explicit lexical meaning provided by hand and arms, they can be divided into two types that are sign and fingerspelling [1,2]. Signs are continuous whereas fingerspelling signs usually change with respect variations in finger angles.

Initially, the algorithms that were designed to provide Sign Language Recognition (SLR) was made possible through data gloves then by advent of digital communication and image processing techniques, vision based systems came into existence that adopted complex data science algorithms. Some of the basic differences between Indian sign language and American Sign Language are shown in Table 1[24].

Sign languages are most important in modern day communication due to increase in man machine interactions and hence it becomes important to design such systems. Most sign languages are not connected with the local spoken languages. Sign supported speech, which combines spoken and sign language and has been used as a basis for learning the grammatical structure of spoken language.

Elakkiya et al,[29] introduced a novel subunit sign modeling framework known as 3-SU for recognizing large vocabulary multimodal signs from continuous video sequences for two different sign languages like German Sign Language (GSL) and American sign Language (ASL). This approach involved modelling with spatial and temporal characteristics of subunits using the Bayesian parallel hidden Markov model (BPaHMM). Their model constructed the sign lexicon for one spatial subunit (hand shape) and two temporal subunits (velocity and position) [29].

Mahmoud et al,[30] proposed a system for recognizing and translating Arabic Sign language. Their system works on the inner circle position which is extracted out by detecting out hand using YCbCr transform on hand and divides the gestures into 4 classes. The features extracted by them were scale invariant, hence their system was more flexible. Their system was able to recognize Arabic letters based on the hand geometry [30].

Shivashankar et al, a classical computer vision approach to recognize a gesture by color transformation from RGB to HSV to YCbCr to Black and White was used. On the obtained processed images thresholding and morphological operations were performed to get a preprocessed image. Then for features it was found roundness of boundary and peaks for recognition [31].

978-1-7281-9744-9/20/$31.00 ©2020 IEEE

Keyframe-Centered Clip (KCC), proposed to realize the high efficiency in isolated Chinese SLR. KCC-level features contain the most representative spatiotemporal information of signs and can be modeled more efficiently than frame-level features [33].

Cui et al[34], employed an approach based on multimodal fusion to be able to integrate both appearance as well as motion details obtained from sign videos that gives better representation related to spatiotemporal details for gestures [34].

Al-Hammadi et al, has stated three different hand gesture database King Saud University Saudi Sign Language (KSU-SSL) dataset, Arabic Sign Language (ArSL) dataset, Purdue RVL-SLLL American Sign Language dataset and classified 40, 23, and 10 classes from these datasets. It was reported that the obtained recognition rates of 98.12%, 100%, and 76.67% on the three datasets, respectively for the signer-dependent mode. For the signer-independent mode, recognition rates of 84.38%, 34.9%, and 70% on these three datasets, respectively [35].

Neel Kamal Bhagat et al, have implemented Convolutional neural networks (CNN) for training 36 static gestures that are related to Indian sign language including alphabets and numbers. The dataset consists of 45,000 RGB images and 45,000 depth images. They have trained the model separately for RGB images and depth images. And they have also trained a single model for both RGB and depth images [36].

Shruti C J et al, used CNN for the classification and recognition. The first step is preprocessing that does the elimination of face to only retain the hand pixels. The method used to eliminate face is Viola-Jones face detection algorithm. After this step the images are used training and testing. The system was successfully trained on all 24 ISL static alphabets with a training accuracy of 99.93% and with testing and validation accuracy of 98.64% [37].

Jasmine Kaur et al, the first step of methodology was pre-processing that includes two steps Skin Color segmentation and Morphological Filtering. They have used ABC (Artificial bee colony) optimization algorithm and FBPNN (Flexible back propagation neural network) is used as a classifier with SIFT as a descriptor. Recognition is done for both English alphabets and numerals from one to ten. They obtained with an average accuracy of 99.43% [38].

Karishma Dixit et al, proposed methodology that consists of three phases. Training phase, testing phase and recognition phase. The pre-processing has two steps segmentation and filtering, Global thresholding method is used for segmentation. They have done recognition for gestures. The dataset includes 720 images. Combinational parameters of Hu invariant moment and structural shape descriptor, gives another feature vector that has to be extracted from the image. Training is done through Multi-class Support Vector Machine (MSVM). This proposed system attained 96% recognition rate [39].

Muthu Mariappan H Et al,developed a system proficient to recognize the ISL in real time. So the captured image consists of half of the body including the face and the gestures. Extracting the ROI (Region of Interest) is an important thing which is done in the segmentation step. After extracting the hand gesture fuzzy clustering method is used for recognition and classification [40].

Yogeshwar I. Rokade Et al, proposed the recognition for finger spelling of ISL. The input for the system is sign in the form of gesture. There are several phases for recognition in this proposed methodology. Initially, the segmentation step was performed on the image to detect the shape of the gesture. The detected region was then transformed into a binary image. After obtaining a binary image the Euclidean distance transformation was applied on that image. HU's moments along with central moments are used for feature extraction [41].

II. SIGN LANGUAGE RECOGNITION SYSTEM

The human hand is a highly deformable object with many degrees of freedom [3]. To recognize the sign language, the input image has to undergo many processes through which the required information will be collected to identify the gesture. There are several methods devised in each process. Common approach is, image can be acquired by using a fixed camera or by using multiple cameras. After obtaining the image it has to be segmented where only the required information will be collected. After segmenting the image, the features will be extracted from the segmented region in order to recognize the gesture. Then the extracted image studied using the learning algorithms. In gesture recognition it makes use of data science techniques involving generation of models, training and testing the model followed by classification of data. Several difficulties were observed to implement the algorithms at each step and methods, these drawbacks can be overcome by successfully following block diagram as mentioned in Figure 1.

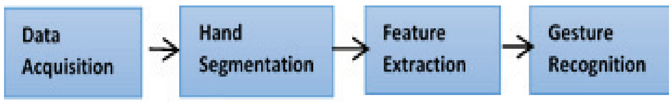

Fig. 1. Basic hand gesture recognition module.

A. Data acquisition

To detect the hand signals data acquisition is the most important step. Segmentation is one of the challenging process to achieve this, one of the earliest method employed was glove based technique, which the signer wears the colored glove, resulting in simplification of the segmentation process. Alternative is a vision-based approach. Image processing algorithms are used in Vision based technique to detect and track hand signs and facial expressions of the signer. This technique is easier to the signer as there is no need to wear any extra hardware as mentioned earlier. Image acquisition is the process of sensing of an image. So in an image acquisition, image is sensed by illumination. It also involves pre-processing steps. In most systems depicted here have used cameras for image acquisition [5], [6].

Glove-based Kinect used to capture the image of the 3D surrounding world, it combines information related to depth and RGB data. As a result, we get a RGBD image of 640x480 resolution, where each pixel gives an information about color and depth [7],[46]. To capture motion signals 8 it makes use of a Kinect sensor that uses time of flight (ToF) techniques to capture depth maps by using an Infrared projector.

Here they used a normal camera of high quality. The images of hand were taken under the same conditions. The background of images is made identical. So, it becomes easier to detect the hand region in the original image by using background subtraction whereas in some cases some moving objects are also included as a result of background subtraction. The skin color is used in discriminating the hand part from mobile parts. The color of the skin is measured with the HSV model. The HSV (hue, saturation, and value) values corresponding to them are H=315, S=94, V=379.

In [10], the data set required was captured with all the background having different color this assisted to remove unsolicited signals, this considers hand signals only, resulting in easier segmentation ToF cameras were employed to capture the images as it helps to remove the depth effect caused due to different illuminations. As well, non-skin color objects are segmented as background by using K-means clustering. Here erosion technique was employed to remove unwanted foreground data.

TABLE I. DIFFERENCES BETWEEN ISL AND ASL

ISL	ASL
Two handed symbols.	One handed symbols.
Symbols are almost similar.	Most symbols are easily distinguishable.
Classification is difficult due to finger angle variations.	Classification is simple and involves only 1or 2 non manual signals.
Sign variations are not easily visible.	Sign variations are easily visualized.

Dutta et al., used a system that extracted the image corresponding to the sign from the video and then eliminated its background using RGB filtering and thresholding. Then they perform edge detection on the binary image and finally extract out posture corresponding to the hand part from it. Then, resize to a standard size and then calculate error using an error matrix for all the images in the database and the processed captured image [11].

After image capturing Dimensionality reduction is of utmost importance this is done using Principal Component analysis (PCA) which also helps in feature extraction [9] where image is reduced to 256x256. Reducing dimension helps in extracting Region of Interest (RoI) and also reduces memory requirement of system [12].

After image capturing and extracting color models, the image is subjected to grey scaling which converts pixel RGB value to a grayscale and reduces data content then it is subjected to filtering to reduce unwanted signals and noise. Image is smoothened by using the mean filter for the complete image and it makes use of BLOB to extract the handout of the image.

In [42] RGB images are retrieved from the camera which comprises various sizes, colors and with different backgrounds to get a better classifier.

In [44] Leap is used for acquiring the images. Leap consists of two cameras and three IR LEDs. It acquires data such as fingertip velocity, range, total finger area, finger length-width ratio and grab and pinch strength which acts as Leap features.

Image uses median filtering to remove salt and pepper noise, it is smoothened using 5x5 Gaussian filter and then it is subjected to morphological operations and finally it is edge detected with sobel operator then it is feature extracted.

OpenPose, an open source toolkit was used in [47] to get the keypoints from a video as the data, which includes 130 keypoints out of which 18 body, 21 each hand 70 face keypoints are given to the feature extractor.

P.K. Athira et al. [13] proposed an appearance based method to recognize dynamic gesture along with a static gesture recognition system they used a video based system with centroid as reference for differentiating gestures, they used co-articulation elimination for gestures. For static gestures they used Zernike moments calculation and used shape based recognition whereas in case of dynamic signs they used a trajectory based classification they considered a 6-dimension feature vector. They also devised an algorithm for co-articulation elimination between static and dynamic gesture combination.

Elpeltagy et al, to obtain best possible accuracy they recorded all the modalities in the data produced by the Kinect sensor. For each sample following modalities were captured: (i) the RGB color video, (ii) the depth video, (iii) the 3D skeleton sequence, (iv) the sequence of hand states (open, closed, or lasso), and (v) the sequence of face features [32].

Neel Kamal Bhagat et al, In their paper they have divided their dataset into two parts. One is for static gestures and another set of dynamic gestures. These images are taken using Microsoft Kinect RGB-D camera [36].

B. Segmentation

The various methods by which the hand segmentation can be performed are mentioned below. Also, the summary of them is shown in Table 2.

1) IR Illumination based segmentation

Here the segmentation of the hand is done using an IR range camera or Time of Flight camera, the camera calculates the distance to various points in an image and clusters the hand region. Therefore, the background elimination using skin color models is not necessary, also it makes the system to be implemented in various places. So it can be said that the IR illumination method is one of most robust system 14. When compared to normal optical localization this system is better with illumination changes 15. But here significant noise is present, and it has to be suppressed by means of median filter operation on the range image [16].

2) Software based

Here it uses software such as LabView for processing image and segmentation. Pre-processing of the images are done with the help of 'Vision Assistant Express VI'. The processing includes background noise suppression and sometimes the images are taken with solid background for easier processing. Later color plane extraction that is conversion to grayscale. And then filtering and thresholding is performed [18]. Hence the segmented image is obtained.

TABLE II. SUMMARY OF SEGMENTATION METHODS

Authors	Methods	Description	Advantages
Neel Kamal Bhagat	Global thresholding algorithm	Thresholding to convert to binary image.	Simple to define threshold value
Pia Breuer Christian Eckes Stefan Müller	IR illumination based segmentation	camera calculates the distance to various points in an image and clusters the hand region.	The background elimination is done at the time of image capturing, hence eliminates the time of finding ROI.
Warrier, Keerthi and Sahu, Jyateen	Software Based	Here it uses software such as LabView	LabView provides various image pre-processing techniques.
Hasan, Mokhtar and Mishra, Pramod.	HSV brightness factor matching	This method is used by deciding the range of H, S, and V parameters.	Simple to define skin HSV values
Chen, Zhi-Hua and Kim, Jung-Tae	BLOB method	The BLOBs are identified	

3) HSV Brightness factor matching

This method is used by deciding the range of H, S, and V parameters. Pigment concentration difference differs the saturation of skin. There are various ethnic groups of color pigment in human skin which all have a different skin color denoted by pigment concentration. Therefore, the color of skin is not a factor in different ethnic groups whereas only HSV range can segment skin from background [17]. In [31], HSV color model is used in detecting the hand from images and the skin pixels that are present in a signed input gesture image.

4) BLOB method

Here in this type of segmentation, the hand region is isolated using the Binary Large OBject (BLOB) concept. After all the pre-processing is done and the image is converted to binary image. Then BLOBs are identified, usually the biggest blobs in an image will be the face and the hand region. But in order to eliminate the face region we use Viola-Jones algorithm to detection of face, then masking the face region and the rest of the image except the hand region, hence the segmented image is obtained.

C. Feature extraction

The various methods by which the features of hand segmented image can be extracted can are mentioned below. Also, the summary of them is shown in Table 3.

1) Edge detection in HSV brightness matching

Edge detection methods are mainly divided into two types, that is gradient and laplacian based which are also called as first and second derivative methods respectively. Here the gradient method is discussed, the edge is maximum value in an image after the first derivative is done. Edge can be easily detected using thresholding techniques, which produces a Binary image (Black and white). This image represents the hand pose, this image is divided into 25 blocks horizontally and vertically creating 625 elements of brightness values. For each image these elements are stored in a database as the feature vectors. Further compression of these feature vectors is done by eliminating redundant 0 value elements. This way HSV values are used for edge detection [17].

2) Hand segmentation using R/G ratio value

First acquire RGB image frame and Separating R, G and B components. If the ratio of R to G is greater than or equal to 1.04 and less than 4 then Convert the skin pixels into white and the rest of them into black i.e. assign the value 255 to skin pixels and 0 to the rest of the pixels else, it is not skin. This skin region detection is kind of unwanted Background cancellation [6].

3) Palm point detection

Palm point is the center of the palm. This is detected using Distance transform. The edges of the palm will have pixel values 0 after the image is converted to a binary image and the hand part will have pixel values [1]. Then using distance transform calculate the longest distance from the 0 pixels that will be the center of the palm [9].

4) Counting fingers using Center of region detection

After detecting the center of region the motive is to find the number of fingers outstretched and their orientation. Then they use the method to find the number of fingers outstretched and their orientation. After obtaining the center of the region, calculate the extreme point of the hand and then calculate the distance from the extreme point of the hand to the center of the region. Once this distance is calculated, draw a circle with the radius equal to 0.7 times the distance from the extreme point to the center of the region. This circle will perfectly intersect all the active fingers in a particular hand gesture. This circle will give information about the number of outstretched fingers. After finding the outstretched fingers we have to find their orientation. Their orientation is found with respect to the wrist line. The range of angles are assigned to each finger. Using this number of outstretched fingers and their orientation we can find the alphabet [6], [19].

5) Dimensionality reduction using LDA

For dimensionality reduction we have different types of machine learning algorithms. One among them is LDA (Linear Discriminant Analysis). Here we first find d-dimensional mean vectors corresponding to each class in the dataset. Then it calculates a scatter matrix both for inter class and intra class. Next we find eigenvectors that are given eigenvalues in the scatter matrix. Then we sort the vectors in accordance with decreasing eigen values to form a d x k matrix W, then we transform into new subspace by using a matrix X which has n x k dimension and then the final matrix is given by Y=X x W which is of dimension n x d and hence reduces the dimensions [4].

6) Gradient Hough Transform (GHT)

Neel Kamal Bhagat at el, in their paper, they have used Gradient Hough Transform (GHT) to obtain different orientation of the hand [36].

7) SIFT technique

Jasmine Kaur at el, used SIFT (scale-invariant feature transform) algorithm to locate the features in the digital image [38].

8) Leap Motion Features

In [44], Leap features are used, this uses feature descriptors such as finger length-width ratio, finger areas, finger range, grab strength of 0 to 1 range being 0 for open hand and 1 for closed hand, pinch strength and also fingertip velocity in millimeters/second.

9) Sub-unit's features

This method of extraction used in [46], various subunits are detected such as location subunits, Hand-arrangement subunits and Motion subunits.

D. Classification and Recognition methods

The most commonly used model for SLR is HMM, it has been used in most of the systems that have been implemented, and some of them have also used CNN and KNN.

1) Hidden Markov Model (HMM)

HMM is a statistical model designed using Bayes network; it is based on probability distribution. HMM was mostly used in speech recognition systems due to its adaptability with infinite hidden features.

Here they design Bayes probability distribution where each state is characterized by its previous state and hence generate a gesture parameter θ that is obtained from output probability densities. Depending on the sequence observed priory distribution on θ. Distribution of θ is assumed to be finite-uniform making P(θ) for any particular θ to be either a constant or zero. The training is done using Baum-Welch form of the expectation-maximization (EM) algorithm, where it computes likelihood of a given sequence due to a particular HMM 20. The testing is done so as to find the sequence which is most likely to produce sequence with similar θ.

TABLE III. SUMMARY OF FEATURE EXTRACTION METHODS

Author	Method	Features	Working
Hasan, Mokhtar and Mishra, Pramod. (2010).	Edge Detection using HSV brightness matching	625 elements of brightness values	Edge is detected using thresholding techniques of gradient or first derivative method
Chen,Zhi-Hua andKim,Jung-TaeandLiang,Jia nningandZhang,J ingandYuan,Yu-Bo. (2014).	Palm Point Detection	Palm points like centre of palm	Detected using distance transform
R.R.Itkarkar and A. V. Nandi	Center of region detection	Finger count and orientation	Distance calculation of extreme points from the center of region
Mahesh Kumar N B, 2018	Dimensionality reduction	Resultant reduced dimension matrix	Linear discriminant analysis is used to find eigenvalues and eigenvectors in the scatter matrix.
Jasmine Kaur et al	SIFT technique	SIFT features of an image	Scale invariant Feature transform is used to get the feature.
Jacob Schioppo Zachary Meyer Diego Fabiano Shaun Canavan, 2019	Leap Motion	Finger length-width ratio, finger areas, finger range, grab strength and fingertip velocity.	Using 2 cameras and 3 IR LEDs in an Egocentric View on HTC Vive Headset

Wang et al., 2006 used Multi-dimensional HMM for American Sign Language (ASL) with 96.7% accuracy, they acquired data from Cyber Glove through the Human computer interface which described features of the gesture. The data was recognized using this stochastic model. Similarly, Liang and Ouhyoung also used HMM in their paper, they acquired data using data glove. The system was designed where used colored images for 10 Arabic numbers and obtained 98.94% of accuracy 21. For Taiwanese Sign Language 22, 23 and Arabic Sign Language 21 used HMM. Comparing the Accuracy of different algorithms, the classical 1D HMM gives accuracy of 85%. Optimized 1D HMM gives accuracy of 92%, classical 2D HMM gives accuracy of 96%, P2D-HMM gives accuracy of 98.5%, and Tcom-P2HMM gives accuracy of 99.5% Table 4.

TABLE IV. SUMMARY OF HMM BASED TECHNIQUES WITH RESULTS

Author	Method(s)	Data	Result (%)
(Wang et al., 2006) 24	Multi-dimensional HMM	26 ASL alphabets and 36 ASL handshapes	96.7
(Dreuw et al., 2007) 25	3 states left-to-right HMM	RWTH-Boston-104 corpus (201 sequence, 104 words)	2.3%(Error rate)
(Kumar et al., 2017) 26	CHMM	25 dynamic words from Indian Sign Language	90.80
(Liang and Ouhyoung, 1998) 23	HMM	250 TSL words	80.4
(Elmezain, Al-Hamadi, Appenrodt, and Michaelis (2008)) 22	HMM	Arabic Numbers(0-9)	98.94% and 95.7% 98.94%(Isolate) 95.7%(Continuous)

2) Artificial Neural Network (ANN)

An artificial neural network is an information processing model originated on the basis of the basic neural structure of the human brain. In biological neurons the input impulse is received by synapses (It is a junction between synaptic neurons and axon). This impulse is carried through the axons. Axons acts a channel to the digital impulse. This impulse is passed to other synapses to reach it to other neurons. This process repeats until impulse reaches the brain.

The same principle is used in artificial neural networks. Artificial neural networks will have a layer, an output layer and one or more middle hidden layers. The input layer will take the desired input and the output will give the output as recognised by the trained hidden layer.

Let us discuss this topic with a clear example. Assume you are playing cricket for the first time and you don't know the rules. Unfortunately, if you throw no ball or wide, you will come to know this is the wrong way of playing and you will correct it for the next time. In the same way artificial neural networks will compare the obtained output with the actual output and calculate the error value and if this error is maximum then it will adjust to the actual output. This is called a backward propagation algorithm which tries to minimize the error until the neural networks completely learns the correct output 27. Yogeshwar I. Rokade at el, they have used many different methods for classification and also compared the results of different methods. They have used SVM (Support vector machine) and neural networks mainly ANN (Artificial neural network).And they have compared the results with different numbers of features 41.

Adithya V, Vinod P. R., and Gopalakrishnan U. (2013) used ANN for classification. Here they made use of four layers, one input and one output with two hidden layers. They made use of a backpropagation algorithm to train the system. They classified 36 ISL signs (26 Alphabets and 10 numbers). For training they used 360 images with 10 for each sign and testing they used 5 images for 36 signs. They were able to get an accuracy 91% from their dataset.

3) k-Nearest Neighbors (kNN)

NN classifiers are no-parametric models. It is based on inter class distance and the decision is based on interclass distance, so this model has ambiguity if the interclass distance is very small and can lead to wrong classification. The main thing to be considered is to what value of k i.e. number of NNs required and it depends on dataset and features and interclass distances. Kausar et al 10 created two groups for classification with the first group having 37 alphabets and second numbers from 1-9. The model worked very well when confusing signs were removed in group 1 but it worked fine for group 2.

4) Support Vector Machine classifier(SVM)

SVM is a classification model that finds optimal hyperplane between different classes using supervised learning on the training data. Given a set of features it checks to which class it belongs. It is one of the best techniques for gesture recognition. P.K. Athira et al have used Multi class SVC classifiers with radial basis function for classification. They were able to get an accuracy of around 90% in their paper [13].

5) Convolutional Neural Network (CNN)

CNN uses convolution layers to extract features. Huang et al. (2015) made use of 3D CNN for sign to text/speech. The accuracy was around 94.2%, they used 25 words of sign language used in daily communications. For continuous SLR they (Koller et al., 2016) made use of a Hybrid CNN-Hidden Markov Model (CNN-HMM) 28. Huang et al, made use of customized spatiotemporal KCC features. For RGB modalities they used CNN, depth motion maps (DMMs) for histogram of oriented gradient (HOG), For Trajectory data used linear sampling to normalize unfixed length trajectory into fixed size then finally fused all of three by concatenating using Long Short-Term Memory (LSTM) network [33].

In [42], [43] they used 100 distinct signs with two 2D convolution layers, a max pooling layer, two dense layers and one dropout and flatten layer and achieved a validation accuracy of 98.70% in [42] and recognition rate of 92.83% in [43].

They have defined CNN architecture with three convolution layers and two max pooling layers.In between them the relu activation layer is used.And after that a 20% dropout layer and two dense layers and a softmax layer is used.

6) Egocentric view Classification

This is a classification method used for Leap features of leap skeletal data using 10-fold cross validation in deep feed forward network with high accuracy. This type of classification is able to recognize dynamic gestures such as J an Z [44].

TABLE V. COMPARISON OF DATABASES USED IN SIGN LANGUAGE RECOGNITION

Name	Country	Classes	Language level	Type
Boston ASL LVD	USA	3300	Word	Video Angles
CUNY ASL	USA	-	Sentence	RGB videos
DEVISIGN-G	China	36 (letters/numbers)	Word	RGB videos
DGS Kinect 40	Germany	40	Word	Video, multiple angles
GSL 20	Greek	20	Word	
IIITA - ROBITA	India	23	Word	RGB videos with 320x240 resolution
LSA64	Argentina	64	Word	RGB videos
MSR Gesture 3D	USA	12	Word	Videos (depth)
PSL Kinect 30	Poland	30	Word	Videos and depth from Kinect camera
RWTH-PHOENIX-Weather	Germany	1200	Sentence	Videos

III. DATABASES

Sign language database is the foundation for sign language recognition in computer vision based technique and can influence the whole recognition system, the resolution and number of images/ videos for signs of different SLR systems are depicted with their characteristics in Table 5.

A. RWTH-PHOENIX-Weather Database

German public TV station used gloss notation for weather forecasting where the transcriptions were done by deaf. The signs have been recorded with a stationary color camera that is placed in front of the sign language interpreters who wore dark clothes with artificial grey background. Videos were captured at 25 fps and with the size of each frame 210 x 260 pixels.

B. Boston ASL LVD

Here they created about 3300 videos from 6 signers for ASL having sign start and end codes for each video they also have numeric ID labels for sign variants. They used 4 cameras with time synchronisation among them 4 cameras capture front view, signers right, face close-up and high resolution front view thus they have different resolutions and frames, after processing all were of 15 fps.

C. DEVISIGN

The DEVISIGN database has around 4400 Chinese sign language vocabularies and involves 330 thousand signs, data from 30 individuals and each data comprises RGB video and also depth and skeleton info. This database provides data having both inter-sign and intra-sign changes. This dataset is again divided into three subsets into G, D and L subsets each of them have eight signers.

D. IIITA -ROBITA Indian Sign Language Gesture Database

This dataset was created in 2009 July IIIT Allahabad at their robotics lab, they captured the videos at 30 fps and having resolution of 320x240 for 23 different gestures and were captured using Sony handycam. Some gestures like above, across, afraid and advances are present which include two handed gestures.

IV. DISCUSSION

The various techniques used at each stage of gesture recognition of sign symbols are shown in Table 6.

TABLE VI. DISCUSSIONS OF VARIOUS TECHNIQUES USED IN GESTURE RECOGNITION SYSTEM

Sl No	Author	Publication	Title of the paper	Signs Used	Segmentation and Feature Extraction Techniques	Classification and Recognition Method	Remarks
1	Sumaira Kausar, M. Younus Javed, Samabia Tehsin, and Almas Anjum.	International Journal of Pattern Recognition and Artificial Intelligence, Vol. 30, No. 5 (2016)	A Novel Mathematical Modeling and Parameterization for Sign Language Classification.	37 PSL alphabets and 9 numbers(1-9)	Centroid estimation and Curve fitting for sign	k-NN classifier	Here they proposed mathematical sign descripted features for classification
2	Honggang Wang, Ming C. Leu and Demil oz.	"Journal Of Information Science and Engineering 22, 1109-1123 (2006)"	American Sign Language Recognition Using Multi-dimensional Hidden Markov Models.	26 alphabets and 36 basic handshapes in the ASL	21 dimensional feature vector with quantization using LBG algorithm	Multi-dimensional HMM	Interactive Training/Learning. The training system adjusts the parameters of the HMM.
3	P.K. Athira, C.J. Sruthi , A. Lijiya	"Journal of King Saud University"	A Signer Independent Sign Language Recognition with Co-articulation Elimination from Live Videos: An Indian Scenario	11 words and 26 ISL alphabets	Region of interest extraction with 6 dimensional feature vector with coarticulation elimination	Training and classification with SVM, Multi class SVC with radial basis function	Keyframe extraction module in this work speeds up the computation and can be used in real time ISL recognition.
4	Mahesh Kumar N B, Sathyamanaglam,	"International Journal of Applied Engineering Research ISSN 0973-4562 Volume 13, Number 9 (2018) pp. 7154-7161.	Conversion of Sign Language into Text	26 ISL alphabets	Otsu algorithm for segmentation with morphological filtering	LDA algorithm for reduction and finds eigen distances for classification	Due to dimensionality reduction the noise will be reduced

5	K. K. Dutta and S. A. S. Bellary *2017*	*"International Conference on Current Trends in Computer, Electrical, Electronics and Communication (CTCEEC)*, Mysore, 2017, pp. 333-336."	Machine Learning Techniques for Indian Sign Language Recognition	25 ISL alphabets(except J)	PCA for dimensionality reduction	Used k-NN and back propagation algorithm for training	Removal of multi-collinearity improves the performance of the machine learning model.

V. CHALLENGES AND OPPORTUNITIES

The prime encounter faced for decoding Sign Lanhuage SLR is the uniformity of Sign Language it is not universal and hence cannot be adapted at universal scale but is limited to regional level. The other challenges that people we may face are related to data acquisition, where system should be able to recognize in all the possible backgrounds. Hence the data acquisition defines how a designed system might operate in future. Other challenges might be related to use of proper algorithm at different stages of operation and their feasibility and efficiencies that require regular testing to design a well-suited system for our objectives.

SLR provides us with an opportunity to better understand sign language and various complexities that are involved with its recognition. To learn various possible techniques that are available, to be able to produce more efficient system for our problem. It also provides us an opportunity to remove communication barrier among people who have problems with speech and are good at gestures to be able to communicate with them without any difficulties. To be able to learn from already available literature and to make use of it for designing a better SLR technique.

VI. CONCLUSION

This paper explains the methods of Sign language recognition and describes the steps involved in gesture recognition which include acquisition, segmentation, feature extraction till recognition and classification. Performance of different models are reviewed and the approaches used by them are explained here. There have been many novel techniques that have been devised for recognition. There have been a lot of efforts to counteract various artifacts that we come across. The systems discussed here have good recognition rates. So as to provide a real time system they have devised many systems with elimination and reduction techniques.

REFERENCES

[1] C.W. Ong and S. Ranganath, "Automatic Sign Language Analysis: A Survey and the Future beyond Lexical Meaning," IEEE Trans. on Pattern Analysis and Machine Intelligence, vol. 27, no. 6, 2005, pp. 873- 891.

[2] Lars Bretzner, Ivan Laptev and Tony Lindeberg, "Hand Gesture Recognition using Multi-Scale Color Features, Hierarchical Models and Particle Filtering" Shortened version in Proc. Face and Gesture 2002, Washington DC, 423–428.

[3] Suharjito, Suharjito and Ariesta, Meita and Wiryana, Fanny and Kusuma Negara, I Gede Putra. (2018). A Survey of Hand Gesture Recognition Methods in Sign Language Recognition. Pertanika Journal of Science and Technology. 26. 1659-1675.

[4] Mahesh Kumar N B,"Conversion of Sign Language into Text", in International Journal of Applied Engineering Research ISSN 0973-4562 Volume 13, Number 9 (2018) pp. 7154-7161.

[5] Gao, Wen and Ma, Jiyong and wu, Jiangqin and Wang, Chunli. (2000). Sign language recognition based on HMM/ANN/DP. IJPRAI. 14. 587-602. 10.1142/S0218001400000386.

[6] Itkarkar, R. and Nandi, Anil. (2013). Hand gesture to speech conversion using Matlab. "2013 4th International Conference on Computing, Communications and Networking Technologies, ICCCNT 2013". 1-4. 10.1109/ICCCNT.2013.6726505.

[7] Rajaganapathy, S. and Aravind, B. and Keerthana, B. and Sivagami, M.. (2015). Conversation of Sign Language to Speech with Human Gestures. Procedia Computer Science. 50. 10.1016/j.pr

[8] Arsan, Taner and Ulgen, Oguz. (2015). Sign Language Converter. International Journal of Computer Science and Engineering Survey. 6. 39-51. 10.5121/ijcses.2015.6403.

[9] Chen, Zhi-Hua and Kim, Jung-Tae and Liang, Jianning and Zhang, Jing and Yuan, Yu-Bo. (2014). Real-Time Hand Gesture Recognition Using Finger Segmentation. TheScientificWorldJournal. 2014. 267872. 10.1155/2014/267872.

[10] Kausar, Sumaira and Javed, Muhammad and Tehsin, Samabia and Anjum, Muhammad Almas. (2016). A Novel Mathematical Modeling and Parameterization for Sign Language Classification. International Journal of Pattern Recognition and Artificial Modi, Krishna and More, Amrita. (2013). Translation of Sign Language Finger-Spelling to Text using Image Processing. International Journal of Computer Applications. 77. 32-37. 10.5120/13440-1313.

[11] Dutta, Kusumika and Bellary, Sunny. (2017). Machine Learning Techniques for Indian Sign Language Recognition. 333-336. 10.1109/CTCEEC.2017.8454988.

[12] Pansare, Jayashree and Gawande, Shravan and Ingle, Maya. (2012). Real-Time Static Hand Gesture Recognition for American Sign Language (ASL) in Complex Background. Journal of Signal and Information Processing. 03. 364-367. 10.4236/jsip.2012.33047.

[13] P.K. Athira, C.J. Sruthi and A. Lijiya ,"A Signer Independent Sign Language Recognition with Co-articulation Elimination from Live Videos: An Indian Scenario",Journal of King Saud University - Computer and Information Sciences.

[14] Pia Breuer Christian Eckes Stefan Müller,"Hand Gesture Recognition with a Novel IR Time-of-Flight Range Camera–A Pilot Study",,"International Conference on Computer Vision / Computer Graphics Collaboration Techniques and Applications,MIRAGE 2007": Computer Vision/Computer Graphics Collaboration Techniques pp 247-260.

[15] Li, Zhi and Jarvis, Ray. (2009). Real time hand gesture recognition using a range camera. "Proceedings of Australasian Conference on Robotics and Automation (ACRA)".

[16] P. Vijayalakshmi and M. Aarthi, "Sign language to speech conversion," 2016 International "Conference on Recent Trends in Information Technology (ICRTIT), Chennai, 2016, pp. 1-6".

[17] Hasan, Mokhtar and Mishra, Pramod. (2010). HSV Brightness Factor Matching for Gesture Recognition System. International Journal of Image Processing. 4.

[18] Warrier, Keerthi and Sahu, Jyateen and Halder, Himadri and Koradiya, Rajkumar and Raj, V. (2016). Software based sign language converter. 1777-1780. 10.1109/ICCSP.2016.7754472.

[19] R. R. Itkarkar and A. V. Nandi, "Hand gesture to speech conversion using Matlab," 2013 "Fourth International Conference on Computing,

Communications and Networking Technologies (ICCCNT), Tiruchengode, 2013, pp. 1-4”.

[20] Lee, G. C., Yeh, F.-H., and Hsiao, Y.-H. (2016). Kinect-based taiwanese sign-language recognition system. *Multimedia Tools and Applications*, *75*(1), 261–279.

[21] Sy Bor Wang, A. Quattoni, L. -. Morency, D. Demirdjian and T. Darrell, "Hidden Conditional Random Fields for Gesture Recognition," ,“2006 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'06), New York, NY, USA, 2006”, pp. 1521-1527.

[22] Elmezain, M., Al-Hamadi, A., Appenrodt, J., and Michaelis, B. (2008). A hidden markov model-based continuous gesture recognition system for hand motion trajectory. In *Pattern Recognition, 2008. ICPR 2008.” “19th International Conference on* (pp. 1–4). Tampa, FL, USA”.

[23] Liang, R. H., and Ouhyoung, M. (1998). A real-time continuous gesture recognition system for sign language. In “*Third IEEE International Conference on Automatic Face and Gesture Recognition* (pp. 558–567)”. Nara, Japan.Shrawankar, Urmila and Dixit, Sayli. (2016). Framing Sentences from Sign Language symbols using NLP.

[24] Wang, H., Leu, M. C., and Oz, C. (2006). American Sign Language Recognition Using Multi-dimensional Hidden Markov Models. *Journal of Information Science and Engineering*, *22*(5), 1109–1123.

[25] Dreuw, P., Rybach, D., Deselaers, T., Zahedi, M., and Ney, H. (2007). Speech recognition techniques for a sign language recognition system. In “*Eighth Annual Conference of the International Speech Communication Association* (pp. 2513-2516)”. Antwerp, Belgium.

[26] Kumar, P., Gauba, H., Roy, P. P., and Dogra, D. P. (2017). Coupled hmm-based multi-sensor data fusion for sign language recognition. *Pattern Recognition Letters*, *86*, 1–8.

[27] Adithya, V. and Vinod, P.R. and Gopalakrishnan, Usha. (2013). Artificial neural network based method for Indian sign language recognition. 1080-1085. 10.1109/CICT.2013.6558259.

[28] Koller, O., Zargaran, O., Ney, H., and Bowden, R. (2016). Deep Sign: Hybrid CNN-HMM for Continuous Sign Language Recognition. In *Proceedings of the “British Machine Vision Conference 2016*. New York, UK”.

[29] R., Elakkiya and Selvamani, K.. (2019). Subunit sign modeling framework for continuous sign language recognition. Computers and Electrical Engineering. 74. 379-390. 10.1016/j.compeleceng.2019.02.012.

[30] Abdo, Mahmoud and Hamdy, Alaa and Salem, Sameh and Saad, El-Sayed. (2014). Arabic Sign Language Recognition. International Journal of Computer Applications. 89. 19. 10.5120/15747-4523.

[31] Ss, Shivashankara and S, Dr.Srinath. (2018). American Sign Language Recognition System: An Optimal Approach. International Journal of Image, Graphics and Signal Processing. 10. 10.5815/ijigsp.2018.08.03.

[32] Elpeltagy, Marwa and Abdelwahab, Moataz and Hussein, Mohamed and Shoukry, Amin and shoala, Asmaa and Galal, Moustafa. (2018). Multi-modality based Arabic sign language recognition. IET Computer Vision. 12. 10.1049/iet-cvi.2017.0598.

[33] Huang, Shiliang and Mao, Chensi and Ye, Zhongfu and Tao, Jinxu. (2018). A Novel Chinese Sign Language Recognition Method Based on Keyframe-Centered Clips. IEEE Signal Processing Letters. PP. 1-1. 10.1109/LSP.2018.2797228.

[34] Cui, Runpeng and Liu, Hu and Zhang, Changshui. (2019). A Deep Neural Framework for Continuous Sign Language Recognition by Iterative Training. IEEE Transactions on Multimedia. PP. 1-1. 10.1109/TMM.2018.2889563.

[35] M. Al-Hammadi, G. Muhammad, W. Abdul, M. Alsulaiman, M. A. Bencherif and M. A. Mekhtiche, "Hand Gesture Recognition for Sign Language Using 3DCNN," in IEEE Access, vol. 8, pp. 79491-79509, 2020, doi: 10.1109/ACCESS.2020.2990434.

[36] Neel Kamal Bhagat, Vishnusai V and Rathna G N “Indian Sign Language Gesture Recognition using Image Processing and Deep Learning” 2019 IEEE 978-1-7281-3857-2/19/.

[37] Sruthi C. J and Lijiya A ”Signet: A Deep Learning based Indian Sign Language Recognition System” International Conference on Communication and Signal Processing, April 4-6, 2019, India.

[38] Jasmine Kaur and C. Rama Krishna “An Efficient Indian Sign Language Recognition System using Sift Descriptor” International Journal of Engineering and Advanced Technology (IJEAT) ISSN: 2249 – 8958, Volume-8 Issue-6, August 2019.

[39] Karishma Dixit and Anand Singh Jalal “Automatic Indian Sign Language Recognition System” 883978-1-4673-4529-3/12/$31.00 2012 IEEE.

[40] Muthu Mariappan H and Dr. Gomathi V “Real-Time Recognition of Indian Sign Language” Second International Conference on Computational Intelligence in Data Science (ICCIDS-2019).

[41] Yogeshwar I. Rokade and Prashant M. Jadav “Indian Sign Language Recognition System” DOI: 10.21817/ijet/2017/v9i3/170903S030 Vol 9 No 3S July 2017.

[42] Wadhawan, A., Kumar, P. Deep learning-based sign language recognition system for static signs. *Neural Comput and Applic* (2020). https://doi.org/10.1007/s00521-019-04691-y

[43] Oyedotun, O.K., Khashman, A. Deep learning in vision-based static hand gesture recognition. *Neural Comput and Applic* 28, 3941–3951 (2017). https://doi.org/10.1007/s00521-016-2294-8

[44] Jacob Schioppo, Zachary Meyer, Diego Fabiano, and Shaun Canavan. 2019. Sign Language Recognition: Learning American Sign Language in a Virtual Environment. In Extended Abstracts of the 2019 CHI Conference on Human Factors in Computing Systems (CHI EA ’19). Association for Computing Machinery, New York, NY, USA, Paper LBW1422,1–6. DOI:https://doi.org/10.1145/3290607.3313025

[45] Li, Dongxu and Rodríguez, Cristian and Yu, Xin and Li, Hongdong. (2019). Word-level Deep Sign Language Recognition from Video: A New Large-scale Dataset and Methods Comparison.

[46] Helen Cooper, Eng-Jon Ong, Nicolas Pugeault, and Richard Bowden. 2012. Sign language recognition using sub-units. J. Mach. Learn. Res. 13, 1 (January 2012), 2205–2231.

[47] Sang-Ki Ko, Jae Gi Son, and Hyedong Jung. 2018. Sign language recognition with recurrent neural network using human keypoint detection. In Proceedings of the 2018 Conference on Research in Adaptive and Convergent Systems (RACS ’18). Association for Computing Machinery, New York, NY, USA, 326–328. DOI:https://doi.org/10.1145/3264746.3264805