International Journal of Engineering and Advanced Technology (IJEAT)
ISSN: 2249-8958 (Online), Volume-12 Issue-3, February 2023

Design Challenges in Effective Algorithm Development of Sign Language Recognition System

Sajeena A, T A Shahul Hameed, Sheeba O

Abstract Sign language is the most putative language among hearing impaired people. They use non-verbal form of communication that involves hand gestures, head or body movement or facial expressions. Of these hand gestures is more widely used. Automatic Sign Language Recognition (ASLR) System can be used to convert these non-verbal signs into text or sound so that normal people can identify them without learning the sign language. ASLR employs Image Processing and Artificial Intelligence (AI) algorithms for effective conversion from sign to sound or text. This review unveils various image processing and AI steps involved in the conversion process, bringing out important topologies in the Image acquisition, segmentation, feature extraction, classification and detection process and a comparative analysis among various topologies. Attempts have been made to shed light into adoption of alternate design strategies in segmentation and feature extraction that enhance the performance in a complex environment.

Keywords: Classification, Feature Extraction, Image Acquisition, Image Segmentation, Vision Based Gesture Recognition

I. INTRODUCTION

According to recent statistics from World Health Organization, mute and hearing-impaired people make more than 5% of the world population. They use sign language for communication. Sign language is a complete natural and well-structured language with its own phonology, morphology, syntax and grammar that uses different expressions for communication. But normal people who are not deaf never try to learn sign language. This may lead to the isolation of hearing-impaired people. This isolation can be reduced to a greater extent by converting the gestures by deaf people into text or voice so that normal people can understand them. This can be done by using a Sign Language Recognition (SLR) system, which identifies the sign and convert them into text or sound which the normal people can understand. Face, hands and the whole-body gestures can be used to show the sign. Each sign has its own meaning. However, unlike normal spoken language there is no globally accepted sign language. Every country has its own sign language [1] like British Sign Language (BSL), American Sign Language (ASL), Pakistan Sign Language (PSL), Korean Sign etc. In India we follow Indian Sign Language (ISL). The two widely followed approach in the sign language recognition are glove / device-based system [2] and vision-based system [3]. In glove-based system, the signer has to wear gloves fitted with sensors. The sensors sense the movement and is processed for the result. This approach extracts signer's movements and posture more accurately, but the sensors, connecting wires and processing unit fitted on the gloves makes the system expensive, complex and difficult to wear. In vision based method, images of the gesture are directly captured and are processed for recognition. Vision based method provide a natural environment and freedom to the signer by reducing the complications of wearing a sensor glove. Vision based gesture recognition method can be classified as, appearance based and 3D model based approach. In appearance-based approach features are extracted form visual appearance of images and recognition is done by comparing these features. 3D model-based approach generally attempts to infer the pose of palm and joint angles of hand in 3D Spatial, and convert into 2D projection. The major difficulty faced by vision-based approach is that the accuracy is often affected by noise, illumination conditions, variation of viewpoint, signer color and the presence of a complex background. Sign language recognition by vision-based approach include these basic steps: Image acquisition, preprocessing, segmentation, feature extraction and classification. In acquisition phase the image, or the frame containing the image is captured using still or video camera. The captured image will be subjected to different preprocessing operations to remove the intruded noise, shadow etc. The Region of Interest (ROI) is retained by segmentation which removes the backgrounds and keep only the gesture portion of the image. These gestures are processed for feature extraction using some colour models. These features can be colours, texture, shapes, edges, corners, curvatures etc. Feature extraction encodes related information in a compressed representation and removes less discriminative data that are insignificant in identifying the gesture. The extracted features will undergo classification operation. Features of each gesture are grouped and are used as a database for the recognition of new gestures. Figure 1 shows the basic block diagram of a sign language recognition system.

Manuscript received on 31 January 2023 | Revised Manuscript received on 09 February 2023 | Manuscript Accepted on 15 February 2023 | Manuscript published on 28 February 2023.

*Correspondence Author(s)

Sajeena A*, Department of Electronics and Communication, University of Kerala, Kerala, India. Email: sajeena@tkmce.ac.in, ORCID ID: https://orcid.org/0000-0002-7972-644X

Shahul Hameed T A, Department of Electronics and Communication, TKM College of Engineering, Kollam (Kerala), India. Email: shahulhameed@tkmce.ac.in, ORCID ID: https://orcid.org/0000-0003-1502-9431

Sheeba O, Department of Electronics and Communication, TKM College of Engineering, Kollam (Kerala), India. Email: shb.odattil@gmail.com, ORCID ID: https://orcid.org/0000-0003-2968-5069

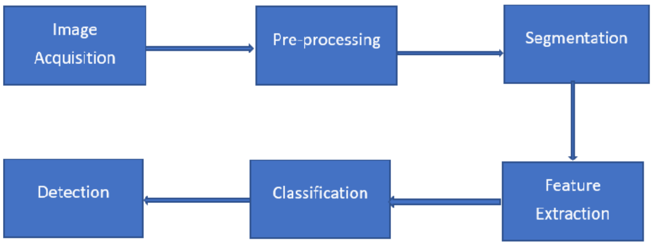

Figure 1: Basic Block Diagram of Sign Language Recognition System

In this paper a comparative study of different processing steps of sign language recognition using vision-based system is done. Here the analysis is done by dividing the overall process into two parts, Image acquisition and preprocessing in the first part and second part include the process of segmentation, feature extraction and classification.

A. Image Acquisition and Pre-processing

Several authors have used camera based acquisition system. The camera can be a mobile camera in [4],[5], [6], [7] which is a hand-held device that normal people cand handle easily. They are also available with different pixel values which make the images with different clarity levels.

Web camera is another choice which most people prefer. Gestures performed in front of a PC is captured by a built-in camera or additional camera fitted with the system [8],[9]. Web camera gives freedom for the signer to perform and the result is directly displayed on the screen. Video cameras can be used to capturing dynamic gestures [10],[11] [1],[12].

Microsoft's Kinect is another popular device used for capturing images [13],[14]. It is a motion sensor developed by Microsoft for X box 360 and Windows PCs for real time gesture recognition. It contains an RGB and a depth sensing camera. Intel RealSense camera is also used by some of the authors [15]. It is a depth sensor used to capture images with information about the distance of the object from the camera.

Acquired images are passed through different preprocessing stages so as to make it suitable for further processing. Resizing, filtering and histogram operations are the usual preprocessing steps.

Median filter is the most popular non linear filter used to reduce noise in the captured images [16]. It smoothen out the image by changing the points with distinct intensity levels to the intensity of their neighbouring pixels.

Morphological operations like opening and closing are also performed by [11], [8], [17], [18] to reduce noise. Open operation involves erosion followed by dilation and is performed to reduce the noise caused by misinterpretation of non-skin pixels as skin pixels. In closing operation erosion follows dilation and is used to reduce errors caused by interpreting skin pixels as non-skin pixels. Morphological operations also smoothen out the images.

One set of authors [19], [20] used Histogram processing as the main preprocessing step. Histogram provides the optimal illumination conditions for capturing an image and thereafter adjust the contrast level for further processing.

Image blurring is also used by some of the authors [21] in preprocessing for noise removal. Blurring removes salt and pepper noise from the image and retains the sharper edges discarding the insignificant low intensity edges.

II. MAJOR WORKS IN IMAGE SEGMENTATION CLASSIFICATION AND DETECTION

Karishma Dixit et al. [16] proposed an Indian Sign Language translator system that uses a Global Thresholding algorithm. Global thresholding can tackle any segmentation problem as a classification problem. A median filter is used in the preprocessing stage to remove noise from the image and feature extraction is done through Hu Invariant Moment which scale and position invariant pattern identification. Hu Invariant Moment (HIM) set is having seven values obtained by normalizing central moments through an order three. HIM is used to scale and position invariant pattern identification even for disjoint shapes which eliminates the use of morphological operators. Seven moments are calculated based on a central moment η_{pq}

$$\eta_{pq} = \frac{\mu_{pq}}{\mu_{pq}^{\Upsilon}}$$

$$\Upsilon = \left[\frac{p+q}{2}\right] + 1$$

$$\mu_{pq} = \sum_x \sum_y (x - \overline{x})^p (y - \overline{y})^q . f(x, y)$$

where, $x = 0,1,2,...., M\text{-}1$; $y = 0,1,.............., M\text{-}1$

$$p, q = 0,1,2,3,...............$$

$$\overline{x} = \frac{m10}{m00}, \quad \overline{y} = \frac{m01}{m00}$$

m00--- Area of subject

m01,m10--- Centre of mass

The image classification is done through a Multi-class Support Vector Machine (MSVM), where each binary classifier is converted to a multiclass classifier and each class receives a unique ID and is stored in a codebook. The test image after preprocessing and segmentation, the features are matched with the code book through MSVM and the most likelihood image is recognized. A recognition rate of 96% is claimed by the authors.

Priyanka C Pankajakshan et al. [11] performed skin colour segmentation based on YCbCr domain on images taken using a video camera with five frames per trigger in RGB domain. The noisy spots resulted due to the variation of light intensity is removed using morphological closing followed by dilation. Feature extraction is done with Canny edge detector which helps in detecting a wide range of edges in the image. 25 images are created for 5 types of gestures and the ANN recognized the gestures.

Shreyashi Narayan Sawant et al. [17] used Otsu algorithm [18] for skin colour segmentation. Otsu's algorithm [22] is a variance-based technique for finding the threshold value where the weighted variance between the foreground and background pixels is the least. The image acquisition done with a web cam is preprocessed through segmentation and morphological filtering [18].

Retrieval Number:100.1/ijeat.C40300212323
DOI: 10.35940/ijeat.C4030.0212323
Journal Website: www.ijeat.org

Published By:
Blue Eyes Intelligence Engineering
and Sciences Publication (BEIESP)

Principal component (PC) is used as the main feature. Principal Component Analysis (PCA) is a dimensionality reduction technique by extracting the desired number of principal components of multidimensional data. Here Eigen values and Eigen vectors are used as the feature components. Minimum Euclidean distance is calculated between the test and train images and the gesture is classified. Eman Thabet et al. [23] performed skin segmentation based on Cb-Cr thresholds either in on-line or off-line mode. Illumination level is avoided to reduce the effect of brightness variations of the scene. In the on-line procedure, Viola-Jones Algorithm is applied to input image frame of the video sequences which is in the YCbCr colour space. Threshold is calculated from the maximum and minimum value of chrominance components. The skin area is segmented through this operation resulting a binary image. Later, Fast Matching Method (FMM) is applied to the segmented features to correct the boundary and compensate for the holes or missing pixels. Offline training is performed under low illumination conditions and face rotation, where Viola–Jones algorithm failed to detect face. G. Ananth Rao et al. [24] proposed edge adaptive thresholding with block variational mean of each 3x3 mask as threshold. The 2D convolution Sobel mask is

$$S^{Mx} = \begin{pmatrix} -1 & -2 & -1 \\ 0 & 0 & 0 \\ -1 & -2 & -1 \end{pmatrix} \quad S^{My} = \begin{pmatrix} -1 & -2 & -1 \\ 0 & 0 & 0 \\ -1 & -2 & -1 \end{pmatrix}$$

The final binary image with a block size **b** $B^x =$

$$\geq p \sum_{i=1}^{b} \sum_{x=1}^{N} \sqrt{(S^{Mx} \otimes \text{ʒ}^{x})^2 + (S^{My} \otimes \text{ʒ}^{y})^2}$$

Gesture video captured using a selfie stick is Gaussian filtered, segmented, morphologically subtracted and subjected to feature extraction. For this Discrete Cosine Transform (DCT) is used along with Principal Component Analysis (PCA). Minimum Distance Classifier is used here since it does not require prior training and uses Mahalanobis distance. Mahalanobis distance includes inter sample covariance's in different directions during distance calculation. It is better than Euclidian distance for sign language classification. Word matching score is calculated as the ratio of correct classification to total number of samples used for classification.

Word Matching Score (WMS) =

$$\frac{Correct\ classification}{Total\ Signs\ in\ a\ VIdeo} \text{ x } 100.$$

The performance of the system is compared with Euclidian distance and normalized Euclidian distance classifiers and the result is shown in Table 1

Table 1: Performance analysis of Euclidian Distance Classifier, Normalized Euclidian Distance Classifier and Mahalanobis Distance Classifier

	Euclidian Distance Classifier	Normalized Euclidian Distance Classifier	Mahalanobis Distance Classifier
Performance (%)	74.11	71.76	90.58
Performance with three testing videos (%)	62.94	61.76	85.88

Oriented FAST and Rotated BRIEF (ORB) feature extraction technique has been tested against different preprocessing techniques such as Histogram of Gradients, LBP and PCA, by Ashish Sharm et al. [19]. ORB [25] uses both FAST key point detector and BRIEF descriptor appeared to be more natural and computationally efficient than LPB and PCA algorithms. K-means clustering is used to reduce the number of features and Canny edge detector that uses a multi stage algorithm to distinguish sharp discontinuities is used in the preprocessing stage to detect the edges.

Table 2: Accuracies of Different Feature Extraction - Classifier Combinations with Varying Number of Test Images

Number of test images	Classifier used	Feature extraction technique	Accuracy
5	SVM	HOG	80
100	SVM	YCbCr- HOG	89.54
800	SVM	SIFT	92.25
17400	KNN	ORB	95.81
17400	MLP	ORB	96.96

Shravani K et al. [26] claims an accuracy of 99% for Indian sign language character recognition for images converted into HSV colour space. They applied a skin mask along with a Canny edge detector for the segmentation of skin pixels. Speeded Up Robust Features (SURF) algorithm which is robust against scaling, rotation, variation and occlusion is used for feature extraction. In SURF algorithm an integral image is generated, which can be used by all the subsequent parts of the algorithm to accelerate their speed. The integral image is given by equation

$$I \sum(x,y) = \sum_{i=0}^{x} \sum_{j=0}^{y} I\ (i,j)$$

A Fast-Hessian detector is used to locate image's significant points. SURF algorithm treats all the significant points with the same weight. The weight of each significant point is given by equation

$$Wp = \frac{No, of\ detected\ images\ w.r.to\ point\ p}{No.of\ training\ images\ in\ object}$$

A number of feature pairs were generated between the test image and the corresponding data set images. These SURF features are clustered with a mini batch K-means which is similar to K-means clustering but with the advantage of improved processing time and memory utilization. The training data is then classified with Support Vector Machine and they obtained an accuracy of 99%. Anup Kumar et al. [10] proposed a method based on skin colour segmentation for static and dynamic gesture recognition. In skin colour segmentation face region will be more prominent than hands. The face region is detected and removed using Viola and Johns algorithm. Viola and Johns is a high-speed real time face detection algorithm using Ada Boost classifier. Skin colour segmentation based on YCbCr colour space on the webcam images is done by Ashish S Niakm et al. [8]. Morphological operations Erosion and Dilation removed the noise from the images in the preprocessing stage.

Contour detection and convexity hull algorithms are used for feature extraction. In the convexity hull algorithm, a bounding rectangle which contain the hull is formed by joining the X- Y coordinates of the palm. It also abstracts the convex defects of the hand which are present in between the valley of two fingers. Average of all such defects give the center of the palm. Finger opening and closing is determined by taking the ratio of palm radius and the distance of hull points from the center of the palm. It is found that Convexity hull algorithm is an appropriate method for finger point detection and number recognition. Hema B N et al. [27] uses Histogram Oriented Gradient (HOG) as the feature descriptor for the images acquired through a web camera. HOG counts the occurrence of gradient orientation in the region of interest of the image. The image portion is divided into small connected regions called cells and a histogram gradient direction is compiled for each pixel within the cell. The concatenation of these histograms forms the descriptor.

Signs are recorded from people who are deaf and mute by birth by Purva C. Badhe [1]. The raw data is initially processed to eliminate redundancy, noise and other information that does not contribute to feature extraction. Then the images are cropped and a difference image is formed by subtracting the consecutive frames. Segmentation is done on YCbCr colour space by setting a proper threshold. Fourier descriptor (FD) is used for feature extraction which calculates the boundary points of hands by the contour following algorithm. 28 feature descriptors for each frame are considered and are compressed before storing into a code book. Compression is achieved through Linde-Buzo-Gray (LBG) which is a lossy compression employing non-uniform many-to-one mapping. In the testing phase, the features of the testing sequence are matched with the reference code book vectors. The distance between these vectors were calculated using Euclidean distance and the code vector which gives the minimum distance of the testing code is considered as the match and the corresponding signer's gesture output is given as the result. Mahesh M et al. [4] proposed a sign language translator for mobile platforms. Device camera capture the images and skin detection is done by combining the result of RGB, YCbCr and HSI methods. The thresholded image is resized to half of the original image before subjected to histogram matching. The images in the data base and the captured images are fed to a comparator which uses histogram matching and ORB descriptor to compare the images and name the gesture if a good match is found.

Here a steered BRIEF is used. The direction from a corner to the centroid is used for steering. A recognition accuracy of 70% is obtained and the main attraction of this work is that the user can add new gestures to the data set.

Geethu G Nath et al. [21] developed a real time ASL language interpreter with an ARM ACOTEX A8 processor on Beagle Bone Processor. Images captured by a webcam were pre-processed through blurring, RGB to binary conversion and edge detection. Canny edge detection is used for binary conversion and Sobel kernel is used for filtering the image. The output pixel value g(i, j) is obtained from the equation

$$g(i, j) = \sum_{(k,I)} f(i+k, j+I)h(k,I)$$

where $f(i+k, j+I)$ is the input pixel value and h(k,I) is the kernel.

The first derivative is obtained in the horizontal and vertical directions and the edge gradient and direction for each pixel is obtained as

Edge gradient, $K = \sqrt{(K_x^2 + K_y^2)}$

Angle, $(\theta) = \tan^{-1}(K_x/K_y)$

The numbers of ASL are recognized by convex hull detection or Jarvis algorithm. The convex and defect points are obtained from the convex hull. It works as envelope around the hand contour. By observing the defect points the number of fingers in the hand sign can be counted. Template matching is used to detect the alphabets. Features of the images stored in the training phase can be used as an aid to hearing impaired people.

Image acquisition by Keerthi S Warrier et al. [5] is done by a smart phone camera with vision acquisition express VI software. Images are taken on solid background and converted to gray scale using colour plane extraction. Linear filtering is done with 'Convolution: Highlight Details' filter which highlights prominent features and a threshold value is applied manually to obtain a binary image. Real time gesture recognition is done the acquired image by applying Geometric Matching Linear Discriminant Analysis (LDA) is used as the classifier by Mahesh Kumar N B [18]. LDA finds a linear combination of features and these features can characterize or separate two or more classes of objects or events. Safar Ahmed Ansari et al. [28] used ISL Dataset consists of 140 static signs to recognize static gestures recorded using a Kinect depth camera. K- means clustering with city- block distance is used as the distance matrix for segmentation. Their algorithm selected the closest clusters on the basis of the clusters mean points depth results a faster clustering. SIFT (Scale Invariant Feature Transform) feature vectors are calculated and indexed them in a k-d tree along with their class labels. Recognition is done using Viewpoint Feature Histogram descriptor (VFH), Speeded-Up Robust Feature (SURF), Neural Networks (NN) and the combination of these three. VFH is a robust view point-invariant descriptor used for extracting features from 3D point clouds. Its computational cost is low and is suitable for real time applications. In SURF the points of interest are calculated by Hessian matrix approach. The second order image intensity variations around a selected area is described by Hessian matrix. SURF based on SIFT uses determinant of Hessian to gauge interest points. They achieved a recognition accuracy of 90% for finger spelling, 100% for three signs and 90.68% for 16 alphabets. SIFT algorithm is used by Sakshi Goyal et al. [29] for feature extraction of Indian Sign Language alphabets. The proposed method uses SIFT keys to identify possible objects in an image in a nearest neighbor approach. The generated feature vectors are invariant to any scaling, rotation or translation of the image.

SIFT algorithm uses four stage filtering for the extraction of these features. The stages are, Scale-Space Extrema detection, Key Point Localization, Orientation Assignment and Key Point Description. The collection of keys that agree on a possible model are identified. The comparison with the highest matched key points in an image will take the lead and will be produced as the output. The proposed algorithm provides 95% accuracy for nine alphabets of Indian Sign Language. Cheok Ming Jin et al. [7] used SURF descriptor with SVM classifier for recognizing 16 gestures of ASL alphabets. The experiment is conducted on a solid background and obtained an accuracy of 97.13%. Same experiment was repeated with SIFT algorithm provided only 92.25% accuracy. Jie Huang et al. [12] used Hierarchical Attention Network with Latent Space (LS-HAN) for continuous sign recognition. The preprocessing steps as well as the problems of temporal segmentation can be eliminated using this method. HAN [30][31]is an extension of Long-Short Term Memory (LSTM). LSTM [32] is an artificial Recurrent Neural Network (RNN). It can process single as well as a sequence (stream) of data. The proposed LS-HAN consists of three components: a two-stream Convolutional Neural Network for video feature representation generation, a Latent Space (LS) for semantic gap bridging, and a Hierarchical Attention Network (HAN) for latent space- based recognition. Latent space captures temporal structures between signing videos and annotated sentences by aligning frames to words.

Yanqiu Liao et al. [14] proposed SLR based on a deep 3-dimensional Residual ConvNet (B3DResNet) and a Bi-directional LSTM network.

The overall procedure is divided into three steps. The first step is the object localization based on Faster R-CNN [33]. R-CNN creates a bunch of bounding boxes, or region proposals, using a Selective Search. These frames are trained by convolution layers for feature extraction. The extracted features are concatenated to get the final feature map. The Region Proposal Network (RPN) slides a small network over the convolutional feature map. RPN provides high quality Region of Interest. B3D ResNet consists of 17 convolutional layers, two Bidirectional-LSTM layers, one fully connected layer, and one soft-max layer. The Bidirectional-LSTM layers analyze the input long term temporal feature sequence and produces an intermediate score. The softmax layer classifies video sequence label and recognizes the dynamic sign language gesture. The video frames can directly be trained in the B3D ResNet model and the show better accuracy then other similar methods.

Table 3: Comparison of HMM-DCT, DNN, C3D, B3D Classification Methods

Methods	Accuracy
HMM-DTC [34]	65.2%
DNN [35]	65.8%
C3D [36]	73.5%
B3D ResNet	86.9%

Siming He [33] compared the object detection algorithms YOLO, Faster R-CNN and Fast R-CNN algorithms. He claimed that Faster R-CNN is more suitable for gesture detection with an accuracy 3% more than Fast R-CNN and 9% than YOLO.

Okan Kopuklu et al. [37] proposed two Convolutional Neural Networks in a hierarchical structure for the classification and recognition of gestures. A light weight CNN (ResNet-10) is used as a detector and a deep CNN (ResNeXt-101) is used for classification. ResNets are deep CNNs with a skip or shortcut connections. These connections make the signals flow easily across the whole network. A sliding window is used for the incoming video stream to feed into the detector. Detector become active only when gestures are being performed and the classifier will respond only when the detector detects a gesture. In the paper they compared the performance of two 3D CNN architectures, C3D and ResNeXt-101. Levenshtein accuracy is used as the evaluation metrics for evaluating two different data sets, EgoGesture and nvGesture. The performance of the classifier for the different data sets is shown on table Table 4

Table 4: Comparison of Res Net and C3D Classification Methods

	Detector's binary classification accuracy	Classifier's classification accuracy	
	ResNet -10	C3D	ResNeXt-101
EgoGesture dataset.	99.68	91.44	94.03
nvGesture dataset	98.02	77.18	83.82

Raw feature and Histogram feature classifiers are used by Tulay Karayilan et.al [38] for recognizing RGB images obtained from Marcel Static Hand Posture dataset. The raw image is resized into 76 x76 px. The resized image flattened to get a 2D array is used as the input to the first classifier and histogram features of the images are given as the input to second classifier. Multilayer Perceptron Neural Network with back propagation algorithm is used for training data and an accuracy of 75% and 85% are obtained respectively for each classifier. Adaptive probabilistic model is used for skin detection by Yogeshwar I. Rokade et al. [39] for Indian Sign Language character recognition. The RGB value of the obtained image is adjusted and is given as the input to the skin detection model. The skin region is detected and the resultant binary image is subjected to morphological operations to eliminate noise. The features are extracted from the distance transform of the binary image. This results another image where each pixel value is replaced by the minimum distance of that pixel from its nearest background pixel. The distance transform d_e is calculated from Euclidean distance as

$$d_e(P,Q)=\sqrt{(x-u)^2+(y-v)^2}$$

where P and Q are the two points.

The summation of the pixel values results a row vector R and a column vector C which represents the shape of the hand. Fourier descriptors of the shape are formed by Fourier transform coefficients of the shape descriptors and is given by

Published By:
Blue Eyes Intelligence Engineering
and Sciences Publication (BEIESP)

$$Vn = \frac{1}{N}\sum_{t=0}^{N-1} R\,(t)\exp\left(\frac{-j2\Pi nt}{N}\right)$$

where n = 0,1,2,3…N-1 and N is size of R.

$$Un = \frac{1}{N}\sum_{t=0}^{N-1} C(t)\exp\left(\frac{-j2\Pi nt}{N}\right)$$

where n = 0,1,2,3…N-1 and N is size of C.

The coefficient of Un and Vn are the fourier descriptors of the shape. The Hu moments are calculated from the geometrical moments of the hand region and the first six Hu moments give the shape of hand. ANN and SVM with a polynomial kernel are the classifiers used. It is concluded that ANN gives higher accuracy than SVM.

A huge data set of 28000 samples were used by Vi N.T.Yruong et.al [40]for classifying static hand gestures of American Sign Language. Images are taken through webcam on complex backgrounds and Adaboost and Haar like classifiers used for classification. An accuracy of 98.7% is claimed by the authors.

Paper [9] authored by Kanchan Dabre used background subtraction, blob analysis, noise reduction, grey scale conversion, brightness and contrast normalization and image scaling as preprocessing steps. Blob analysis can be used for object tracking where it discovers the region of interest (ROI) for further processing. ROI is formed by finding all connective parts of the frame and choose the largest area amongst them. Gaussian filter is used for noise reduction and histogram equalization is performed to normalize brightness and contrast of the frame. The image is resized to 45x45 for further processing. Haar cascaded classifier is used which identifies the region of interest and is analyzed to identify the contrast between the images. Based on this the required cascade is build and the threshold is obtained by analyzing each coordinate of hand sign. The Haar function is given by

$$H(t)\ \begin{cases} 1 & 0 \le t \le ½ \\ -1 & ½ \le t \le 1 \\ 0 & \text{otherwise} \end{cases}$$

Experiment is performed with two distinguished signs and an average accuracy of 92.68% was obtained.

Pichao Wang et al. [3]in their work on continuous gesture recognition segments individual gestures from a depth sequence based of Quality of Movement (QOM) [41].

An Improved Depth Motion Map (IDMM) is constructed based on the absolute difference between the current frame and the start frame for each segment. The resultant IDMM is then converted to pseudo-RGB image and passed on to ConvNet for classification. The pseudo-colour image exploit the texture in the IDMM that corresponds to motion patterns of actions. A Power Rainbow Transform converts these patterns into normalized RGBs as

$$R^* = \{(1+ \cos(\frac{4\pi}{3\ x\ 255}\,I))/2\}^2$$

$$G^* = \{(1+ \cos(\frac{4\pi}{3\ x\ 255}\,I\ - \frac{2\pi}{3}))/2\}^2$$

$$B^* = \{(1+ \cos(\frac{4\pi}{3\ x\ 255}\,I\ - \frac{4\pi}{3}))/2\}^2$$

Where I is the given intensity and R*, G* and B* are the normalized RGB values. The IDMMs are resized into 256 x 256 before feeding into ConvNet. The performance is evaluated as Mean Jaccard Index and a value of 0.2655 is obtained. Images acquired by Intel RealSense camera is used to translate American Sign Language by Jayan Mistry et al. [15]. A rotation quaternion of 20 joints of the hand, the degree of flexion of each finger, degree of openness of each hand and a palm orientation quaternion are the features considered for detection. A StandardScaler rescales each feature independently so as to obtain a mean value zero and a standard deviation of one. The features are scaled to a range between zero and one by a MinMax Scaler and divide them by its maximum absolute value by a MaxAbsScaler Finally the median is brought to zero by RobustScaler followed by scaling the data according to the interquartile range. This will make it more robust to outliers than a StandardScaler Principal Component Analysis (PCA) is used for normalization. In the recognition phase Support Vector Machine (SVM) and Multilayer Perceptron (MLP) were used and their performance were compared.

Table 5: Comparison of Accuracies of Multilayer

Pre-Processing Technique	Accuracy of Multilayer Perceptron (Percentage)	Accuracy of Support Vector Machine (Percentage)
No Preprocessing	81.5	86.1
Standard Scaler	83.4	89.6
Min Maxscaler	84.8	88.2
Max Abs Scaler	92.1	95.0
Robust Scaler	45.0	47.0
Normalization	82.1	87.4

Perceptron and SVM for different preprocessing techniques

It is found that Support Vector Machine with pre-processing by a MaxAbScaler yields the best result. Varunkumar et al. [43] propose Denoising sparse Autoencoders for feature extraction and Softmax classifier for classification of hand gestures. Autoencoders are simple learning circuits that produces outputs with least possible amount of distortion [44]. The cost function used is

$$J(W,b) = \left[\frac{1}{m}\sum_{i=1}^{m}\left(\frac{1}{2}||h_{W,b}(x^{(i)}) - y^{(i)}||^2\right)\right] + \frac{\lambda}{2}\sum_{l=1}^{n_l-1}\sum_{i=1}^{s_l}\sum_{j=1}^{s_l+1}\left(W_{ij}^{(l)}\right)^2$$

Where W, b are the network parameters weight and bias

$x^{(i)}$ is the i^{th} input vector

$y^{(i)} = x^{\prime(i)} = x^{(i)}$, n_i is the number of layers

$h_{w,b}(x^{(i)})$ is the hypothesis or observed output when input is $x^{(i)}$

λ is the weight decay (regularization) parameter used to prevent overfitting.

Autoencoder tries to learn an approximation to identity function. Additional constraints can be placed on the network so as to limit the number of hidden units to be less than the number of input units.

Published By:
Blue Eyes Intelligence Engineering
and Sciences Publication (BEIESP)

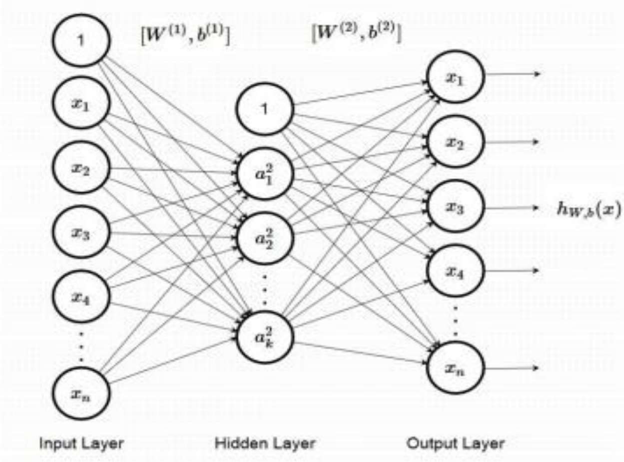

Figure 2: Autoencoder

Denoising autoencoders have hidden units greater than or equal to the number of input units and a sparsity constraint is added to the hidden units to make it a Sparse Autoencoder. A sparsity parameter ρ is added which make the hidden unit to be zero most of the time whenever the activation function taken is sigmoid function. The overall cost function used here is

$$J_{sparse}(W,b) = J(W,b) + \beta \sum_{j=1}^{s_2} KL\left(\rho || \hat{\rho_j}\right)$$

The features extracted through these layers are passed through a Softmax classifier which is added as the last layer to classify the image. Finally, the cost function is calculated using the L-BFGS algorithm [45].

The technique is used to identify 20 gestures of American Sign Language and a classification accuracy of 83.36% is obtained.

III. COMPARISON OF RESULTS

Automatic sign language recognition system helps to remove the communication barrier between the normal and the hearing impaired people. SLR has become one of the fast developing field in the human computer interaction where researchers has to face so many challenges. The various issues related with image-based sign language recognition are:

1. Image acquisition: Images can be acquired through web cam [11],[8],[17],[38],[40],[9],[26], mobile camera [4],[5],[6],[7] Kinect sensor [47],[13],[14],[48],[28], Intel RealSense camera [20],[42] and video camera [10],[11],[1],[12]. Input images are also obtained through datasets provided by different agencies [38],[49],[18],[50],[51],[14],[43],[39]. Processing of images acquired in real time may have problems like orientation of camera, lighting conditions, background objects, color of signer etc.

2. Segmentation: Segmentation is obtained through two main techniques, segmentation based on external aids like colour gloves and based on skin colour. In this review the first method is not considered. Skin colour segmentation can be implemented based on colour schemes like RGB [38],[4],[21],[6],[7],[33], HSV [10], [26], Cb-Cr [23] and YCbCr [11],[8],[1],[4]. Of these YCbCr is consistent with human visual perception and more often used for skin colour modelling. Other algorithms used for skin segmentation are Morphological filtering QOM [3], LS HANS continuous SLR [12], Faster RCNN [12], Sobel edge operation morphology [24], edge adaptive thresholding [24], Canny edge detection [48], City-block distance [28] and adaptive probabilistic model [39]. Background colour/objects, lighting conditions and skin colour variations are the main challenges associated with skin colour segmentation.

3. Feature extraction: Feature extraction provides most relevant information from the processed image so as to make the classifier work efficiently with limited amount of training data. Canny edge detector [11], [5], [26], Convex Hull [8], [21], Principle Component Analysis (PCA) [17], [46], [11], [18], [15], [24], Scale Invariant Fourier Transform (SIFT) [29], [28], Speeded Up Robust Features (SURF) [7], [26], Discrete Cosine Transform (DCT) [24] are some of the algorithms we reviewed in this survey. Selection of proper feature extraction method is a major challenge in gesture recognition since the selected feature may greatly influenced by the position and orientation of hand, way of expression of the sign, and the proceeding and succeeding signs.

4.Classification and recognition: The above-mentioned difficulties with feature extraction can be greatly overcome by selecting a proper classification technique. Support Vector Machine (SVM) [10], [16], [6], [7], [42], [26], [39], ANN [11], [38], [27], [39], R-CNN [33], Oriented Fast and Rotated Brief (ORB) [4], K-Nearest Neighbor (KNN) [2],[46], Haar [40], [9], Softmax [43], ConvNet [3], ResNet [14], Hierarchical Attention Network (HAN) [12] are the classification algorithms viewed here. The selection of proper algorithm is determined by static/ dynamic sign, speed of action, repeated signs, manual and non-manual signs etc.

For recognizing a sign, the image has to pass through all these stages. Different combinations of algorithms achieve different accuracy levels as shown in Table 6

IV. CONCLUSION

People with hearing and speaking impairment cannot exist without sign language. A gesture recognition system is a human machine interactive system which helps the normal people to interact with hearing impaired people. Most of the research works were done with static gestures only and a very few are done with dynamic gestures. The reviewed papers that uses different approaches on various steps of SLR. Image acquisition can be done with mobile/ webcam, Kinect or RealSense camera. Kinect and RealSense camera provides good acquisition rate and quality but they are difficult to use in public places and are costlier too. Varying skin tone is the main challenge in the segmentation phase especially in Indian Sign Language recognition system. Attempts to include facial expression along with hand gesture makes the system more complicated. Different algorithms used for feature extraction and classification were also discussed. But the comparison of these methods is subjective, since each method has its own strength and limitations compared to other methods.

Retrieval Number:100.1/ijeat.C40300212323
DOI: 10.35940/ijeat.C4030.0212323
Journal Website: www.ijeat.org

Published By:
Blue Eyes Intelligence Engineering
and Sciences Publication (BEIESP)

Different combinations of these methods yield results with varying accuracies. Lack of standardized data set is another difficulty that the researchers are facing which forces them to use small vocabulary with self-made data set. All these things make the research confined to the signs of a particular country or region.

Table 6: Comparison Table- Segmentation, Feature Extraction and Classification Methods

Ref. No.	Author	I/P	Segmentation	Feature Extraction	Classification/ Recognition	Accuracy
1	'Indian Sign Language Translator Using Gesture Recognition Algorithm', Purva C. Badhe	Video	YCbCr	2D FFT Fourier Descriptors	Code book generated with Linde- Buzo- Gray (LBG) type vector quantization.	97.50%
2	R. Martin McGuire1, 'Towards a One-Way American Sign Language Translator',Jose Hernandez-Rebollar2, Thad Starner	Data Glove			KNearest Neighbors and Convolutional Neural Networks	97.80%
3	Large-scale Continuous Gesture Recognition Using Neighbors and Convolutional Neural Networks',Pichao Wang, Wanqing Li,	ChaLearn LAP ConGD Dataset	quantity of movement (QOM)	ConvNets	ConvNet	0.2655 (Mean Jaccard Index)
4	Mohammed Elmahgiubi,'Sign Language Translator and Gesture Recognition,	Sensor gloves	Otsu algorithm, Morphological Filtering	Principal component analysis		
5	'Software Based Sign Language Converter', Keerthi S Warrier	Smart Phone Camera with vision acquisition express VI software	Gray scale	Edge detection	Template matching	
6	Zulfiqar A. Memon, 'Real Time Translator for Sign Languages'	Mobile camera	RGB		SVM	
7	Cheok Ming Jin, Zaid OmarA Mobile Application of American Sign Language Translation via Image Processing Algorithms	Mobile camera	RGB	Speeded Up Robust Features (SURF),SIFT	Kmeans Clustering,Support Vector Machine (SVM)	97.13
9	Machine Learning Model for Sign Language Interpretation using webcam images',Kanchan Dabre	Web cam	Gray scale		Haar Cascade Classifier	92.68%.
10	Sign language recognition - Anup Kumar	Video	Skin Colour, HSV	Zernike moments and curve feature	Multiclass SVM	>90
11	Sign Language Recognition - System, Priyanka C Pankajakshan	Video/Webcam	Skin Colour, YCbCr	Canny Edge detector	ANN	
12	Video-based Sign Language Recognition without Temporal Segmentation, Jie Huang1,	Video clips	LS-HAN framework for continuous SLR	Two strean 3D CNN	Hierarchical Attention Network , (HAN) for continuous SLR in a latent space.	82.7
13	Conversation of Sign Language to Speech with Human Gestures, Rajaganapathy.	Microsoft Kinect		Position of 20 joints		90

Retrieval Number:100.1/ijeat.C40300212323
DOI: 10.35940/ijeat.C4030.0212323
Journal Website: www.ijeat.org

Published By:
Blue Eyes Intelligence Engineering and Sciences Publication (BEIESP)

15	Indian Sign Language Translator using the Intel RealSense Camera,Jayan Mistry	Intel RealSense F200 camera and the RealSense API		Principal Component Analysis (PCA)	support vector machine (SVM) and a multilayer perceptron (MLP)	95.0%, with an SVM and the MaxAbsScaler (with 92.1%)
16	Automatic Indian Sign Language Recognition System, Karishma Dixit	Image	Global thresholding algorithm	Structural Shape Descriptors	Multiclass SVM	96.23%
17	Real Time Sign LanguageRecognition using PCA', Shreyashi Narayan Sawant,	Web cam	Otsu algorithm	Principle Component Analysis	Euclidean distance	
18	'Conversion of Sign Language into Text', Mahesh Kumar N B	From database	Otsu algorithm	Morphological Filtering + Principle component analysis	Linear Discriminant Analysis (LDA) for recognition	
21	Real Time Sign Language Interpreter',Geethu G Nath, Arun C S,	USB camers	RGB	Convex hull or Jarvis algorithm	Template matching	
22	Real Time Hand Gesture Recognition Using Different Algorithm Based On American Sign Language', Prof. B.B.Gite	camera	Canny edge detection and otsu's techniques		CNN	
23	Low Cost Skin Segmentation Scheme in Videos Using Two Alternative Methods for Dynamic Hand Gesture Detection', Eman Thabetod	Data set	Skin colour-Cb-Cr Thresholds	Fast Matching Method (FMM)		
24	Selfie video based continuous Indian sign language recognition system', G Ananth Rao, P V V Kishore,	Smart phone front camera.	Sobel edge operator, morphology and Edge adaptive thresholding	Discrete Cosine Transform (DCT) along with Principle Component Analysis (PCA).	minimum distance classifier (MDC).	WMS-85.58 % ANN-90%
26	'Indian Sign Language Character Recognition', Shravani K	Webcam	HSV	Canny Edge Detector, SURF features trained in mini batch K means	Bag of visual words (BoW), SVM	99%
27	Sign Language ad Gesture Recognition for Deaf and Dump People', Hema B N, Sania Anjum	Web camera	Binary	histogram of oriented gradients (HOG)	ANN	
28	'Nearest neighbour classification of Indian sign language gestures using kinect camera', Zafar Ahamed Ansari	Kinect Depth Camera	City block distance	Scale Invariance Fourier Transform (SIFT)	Viewpoint Feature Histogram descriptor (VFH), SURF, NN	90% for finger spelling, 100% for three signs and 90.68% for 16 alphabets
29	'Sign Language Recognition System For Deaf And Dumb People', Sakshi Goyal, Ishita Sharma	Integrated webcam		Key point detection - Scale Invariance Fourier Transform (SIFT)		
33	Research of a Sign Language Translation System Based on Deep Learning' Siming He		RGB	3D CNN LSTM	Faster R-CNN	99
38	Sign language recognition', Tulay Karayilan	Web cam Marcel Static Hand Posture (dataset)	RGB	Normalized numerical array	ANN (Raw feature and Histogram Feature) Back propagation algorithm	70% 85%
39	Indian Sign Language Recognition System' , Yogeshwar I. Rokade	Dataset	Skin detection using adaptive probabilistic model	Distance transform of binary image	ANN, SVM	ANN-94.37% SVM-92.12%

Retrieval Number:100.1/ijeat.C40300212323
DOI: 10.35940/ijeat.C4030.0212323
Journal Website: www.ijeat.org

Published By:
Blue Eyes Intelligence Engineering
and Sciences Publication (BEIESP)

43	'Static Hand Gesture Recognition using Stacked Denoising Sparse Autoencoders', Varun Kumar, G.C. Nandi	Standard ASL dataset		Denoising Sparce Autoencoder	Softmax classifier	83.36
46	'American Sign Language Interpreter System for Deaf and Dumb Individuals', Sruthi Upendran		Gray scale	Principle Component Analysis	k- Nearest Neighbor	
51	'Dynamic Sign Language Recognition Based on Video Sequence with BLSTM-3D Residual Networks', Yanqiu Liao	Kinect camera, DEVISIGN_D dataset and SLR_dataset	FAster R CNN	B3D ResNet	B3D Res Net	DEVISIGN-D dataset and SLR Dataset, 89.8 % and 86.9% separately

DECLARATION

Funding/ Grants/ Financial Support	No, I did not receive.
Conflicts of Interest/ Competing Interests	No conflicts of interest to the best of our knowledge.
Ethical Approval and Consent to Participate	No, the article does not require ethical approval and consent to participate with evidence.
Availability of Data and Material/ Data Access Statement	Not relevant.
Authors Contributions	All authors have equal participation in this article.

REFERENCES

1. Purva C Badhe, Vaishali Kulkarni, "Indian Sign Language Translator Using Gesture Recognition Algorithm", International Conference on Computer Graphics, Vision and Information Security, IEEE, 2015 [CrossRef]
2. R. Martin McGuire, Jose Hernandez-Rebollar, Thad Starner1, Valerie Henderson1, Helene Brashear1, and Danielle S. Ross," Towards a One-Way American Sign Language Translator", e Sixth IEEE International Conference on Automatic Face and Gesture Recognition, 2014
3. Pichao Wang, Wanqing Li, Song Liu, Yuyao Zhang, Zhimin Gao and Philip Ogunbona, " Large-scale Continuous Gesture Recognition Using Convolutional Neural Networks", 23rd International Conference on Pattern Recognition, , IEEE Xplore, April 2017
4. Mohammed Elmahgiubi, Mohamed Ennajar, Nabil Drawil, and Mohamed Samir Elbuni, "Sign Language Translator and Gesture Recognition", Global Summit on Computer and Information Technology, IEEE Xplore, Dec 2015, [CrossRef]
5. Keerthi S Warrier, JyateenKumar Sahu, Himadri Halder, Rajkumar Koradiya, and Karthik Raj V," Software Based Sign Language Converter", International Conference on Communication and Signal Processing, , IEEE Xplore, Nov 2016 [CrossRef]
6. Zulfiqar A. Memon, M. Uzair Ahmed, S. Talha Hussain, Zahid Abbas Baig, Umer Aziz, "Real Time Translator for Sign Languages", International Conference on Frontiers of Information Technology, 2017
7. Cheok Ming Jin, Zaid Omar, Mohamed Hisham Jaward, " A Mobile Application of American Sign Language Translation via Image Processing Algorithms", IEEE Xplore, 25 July 2016
8. Ashish S Nikam, Aarti G. Ambekar, "Sign Language Recognition using Image Based Hand Gesture Recognition Techniques", International Conference on Green Engineering and Technologies, IEEE, 2016
9. Kanchan Dabre, Surekha Dholay, "Machine Learning Model for Sign Language Interpretation using webcam images", International conference on Circuit, Systems, Communication and Information Technology, IEEE Xplore, Jun 2014 [CrossRef]
10. Anup Kumar, Karun Thankachan, Mevin M Dominic, "Sign Language Recognition", 3rd International conference o Recent Advances in Information Technology, IEEE Explore, 09 July 2016 [CrossRef]
11. Priyanka C Pankajakshan, Thilagavathy, " Sign Language Recognition System", IEEE sponsored 2nd International Conference on Innovations in Information and Embedded and Communication Systems, 2015 [CrossRef]
12. Jie Huang , Wengang Zhou , Qilin Zhang , Houqiang Li , Weiping Li, "Video-based Sign Language Recognition without Temporal Segmentation", The Thirty-Second AAAI Conference on Artificial Intelligence, Jan 2018 [CrossRef]
13. Rajaganapathy. S1 , Aravind. B, Keerthana. B , Sivagami. M, "Conversation of Sign Language to Speech with Human Gesture", ScienceDirect, Procedia Computer Science, Volume 50, 2015, Pages 10-15 [CrossRef]
14. Yanqiu Liao, Pengwen Xiong, Weidong Min, Weiqiong Min, Jiaho Lu, " Dynamic Sign Language Recognition Based on Video Sequence with BLSTM-3D Residual Networks", IEEE Access, March 2019.
15. Jayan Mistry, Benjamin Inden " An Approach to Sign Language Translation using the Intel RealSense Camera", 10th Computer Science and Electronic Engineering Conference - University of Essex, Colchester, 19-21 September 2018 [CrossRef]
16. Karishma Dixit, Anand Singh Jalal, "Automatic Indian Sign Language Recognition System" 3rd International Advance Computing Conference,2013.
17. Shreyashi Narayan Sawanth, M.S Kumbhar, " Real Time Sign Language Recognition using PCA", IEEE International on Advanced Communication Control and Computing technologies, 2014
18. Mahesh Kumar N B, "conversion of Sign Language into Text", International Journal of Applied Engineering Research ISSN 0973-4562 Volume 13, pp. 7154 7161, Number 9, 2018
19. Asish Sharma, Anmol Mithal, Savitoj Singh, Vasudev Awatramani, "Hand Gesture Recognition using Image Processing and Feature Extraction Techniques", International Conference on Smart Sustainable Intelligent Computing and Applications, 2020 [CrossRef]
20. Md. Shahinur Alam, Ki-Chul Kwon , and Nam Kim, " Md. Shahinur Alam , Ki-Chul Kwon , and Nam Kim", IEEE Transactions on Human- Machine Systems, Vol.51, No. 3, June 2021.
21. Geethu G Nath, Arun C S, "Real Time Sign Language Interpreter", International Conference on Electrical, Instrumentation and Communication, IEEE Xplore, Dec 2017
22. Prof. B B Gite, Vaibhavi Honmane, Vaibhavi Joshi, Rutuja Waghe, "Real Time Hand Gesture Recognition Using Different Algorithm Based On American Sign Language", International Journal of Future Generation Communication and Networking Vol. 13, No.3s, pp. 1031–1036, 2020
23. Eman Thabet, Fatimah Khalid, Puteri Suhaiza Sulaiman, and Razali Yaakob, " Low Cost Skin Segmentation Scheme in Videos Using Two Alternative Methods for Dynamic Hand Gesture Detection Method", Research Article, Hindawi Advances in Multimedia Volume 2017, Article ID 7645189, 9 pages, 2017 [CrossRef]
24. G Ananth Rao, P V V Kishore, " Selfie video based continuous Indian sign language recognition system", Ain Shams Engineering Journal 9, 1929–1939, 2018 [CrossRef]

25. Prashant Aglave , Vijaykumar.S. Kolkure, “Implementation Of High Performance Feature Extraction Method Using Oriented Fast And Rotated Brief Algorithm”, International Journal of Research in Engineering and Technology, Volume: 04 Issue: 02,Feb-2015 [CrossRef]
26. Shravani K, Sree Lakshmi A, Sri GeethikaM, Dr. Sapna B Kulkarni, “ Indian Sign Language Character Recognition”, Journal of Computer Engineering, Volume 22, Issue 3, Ser. I PP 14-19, May - June 2020.
27. Hema B N, Saina Anjum, Umme Hani, Vanaja P, Akshatha M, “Sign Language ad Gesture Recognition for Deaf and Dump People”, International Research Journal of Engineering and Technology, IRJET, Volume 6, Issue 3, March 2019
28. Zafar Ahmed Ansari, Gaurav Harit, “Nearest Neighbour Classification of Indian Sign Language Gestures”, Indian Academy of Sciences, Vol.41, pp. 161-182, February 2016 [CrossRef]
29. Sakshi Goyal1 , Ishita Sharma2 , Shanu Sharma, “ Sign Language Recognition System For Deaf And Dumb People”, International Journal of Engineering Research & Technology, Vol. 2 Issue 4, April 2013.
30. Ebrahim Karami, Siva Prasad, and Mohamed Shehata, “ Image Matching Using SIFT, SURF, BRIEF and ORB: Performance Comparison for Distorted Images”,
31. Yingwei Pan, Tao Mei, Ting Ya, Houqiang Li, and Yong Rui, “ Jointly Modeling Embedding and Translation to Bridge Video and Language”. arXiv preprint arXiv:1505.01861. 2015
32. Liu, T.; Zhou, W.; and Li, H, “Sign language recognition with long short-term memory”, IEEE International Conference on Image Processing, 2871–2875, 2016 [CrossRef]
33. Siming He, “Research of a Sign Language Translation System Based on Deep Learning”, Research of a Sign Language Translation System Based on Deep Learning,2019
34. H. Wang, X. Chai, and X. Chen, “Sparse observation (so) alignment for sign language recognition,” *Neurocomputing*, vol. 175, no. 29, pp. 674-685, Jan. 2016. [CrossRef]
35. T. Kim *et al.*, “Lexicon-free fingerspelling recognition from video: data, models, and signer adaptation,” Computer Speech& Language, vol. 46, pp. 209-232, Nov. 2017. [CrossRef]
36. T. Liu, W. Zhou, and H. Li, “Sign language recognition with long short-term memory,” *IEEE Int. Conf. Image Process.(ICIP)*, Phoenix, AZ, USA, Sep. 2016, pp. 2871- 2875. [CrossRef]
37. Okan Kopuku, Ahmet Gunduz, Neslohan, Kose, Gerhard Rigoll, “ Real-time Hand Gesture Detection and Classification Using Convolutional Neural Networks”,
38. Tulay Karayilan, Ozkan Kilic, “ Sign Language Recognition”, 2nd International Conference on computer Science and Engineering, 2017 IEEE [CrossRef]
39. Yogeshwar I. Rokade , Prashant Jadav, “Indian Sign Language Recognition System” ,International Journal of Engineering and Technology · July 2017, 9(3S):189-196 [CrossRef]
40. Vi N.T Truong, Chuan-Kai Yang, Quoc-Viet Tran, “ A Translator for American Sign Language to Text and Speech”, 5th IEEE Global Conference on Consumer Electronics, IEEE Xplore, Dec 2016 [CrossRef]
41. F. Jiang, S. Zhang, S. Wu, Y. Gao, and D. Zhao, “Multi-layered gesture recognition with kinect,” Journal of Machine Learning Research, vol. 16, no. 2, pp. 227– 254, 2015.
42. Jie Huang; Wengang Zhou; Houqiang Li; Weiping Li “Sign language recognition using real-sense”, 2015 IEEE China Summit and International Conference on Signal and Information Processing IEEE Xplore: September 2015 [CrossRef]
43. Varun Kumar, G.C. Nandi, ‘Static Hand Gesture Recognition using Stacked Denoising Sparse Autoencoders’, 2014 Seventh International Conference on Contemporary Computing, IEEE Xplore. [CrossRef]
44. Pierre Baldi. Autoencoders, unsupervised learning, and deep architectures. In ICML Unsupervised and Transfer Learning, pages 37–50, 2012.
45. Jiquan Ngiam, Adam Coates, Ahbik Lahiri, Bobby Prochnow, Quoc V Le, and Andrew Y Ng. On optimization methods for deep learning. In Proceedings of the 28th International Conference on Machine Learning (ICML-11), pages 265–272, 2011
46. Sruthi Upendran, Thamizharasi. A, “American Sign Language Interpreter System for Deaf and Dumb Individuals”, 2014 International Conference on Control, Instrumentation, Communication and Computational Technologies, July 2014 [CrossRef]
47. Kin Fun Li, Kylee Lothrop, Ethan Gill, and Stephen Lau, “A Web-Based Sign Language Translator Using 3D Video Processing”, The 14th International Conference on Network- Based Information Systems, NBiS 2011, Tirana,Albania, September 7-9,2011
48. Asma Ben Hadj Mohamed, Amal Baghdadi1 , Thierry Val, Laurent Andrieux Abdennaceur kachouri, “Edges detection in depth images for a gesture recognition application using a Kinect WSN”, 5th International Conference on Web and Information Technologies (ICWIT 2013), 9 May 2013
49. Dimitrios Konstantinidis, Kosmas Dimitropoulos and Petros Daras, ‘SignLanguage Recognition based on Hand and Body Skeletal Data’, 3D TV conference,IEEE Xplore, October 2018 [CrossRef]
50. Amanda Duarte, “Cross-modal Neural Sign Language Translation”, Doctoral Symposium, MM ’19, October 21–25, Nice, France, 2019. [CrossRef]
51. Yanqiu Liao, Pengwen Xiong, Weidong Min, Weiqiong Min, JIahao, “Dynamic Sign Language Recognition Based on Video Sequence with BLSTM-3D Residual Networks”, IEEE Access, 2019

AUTHORS PROFILE

Sajeena A, completed her Bachelor’s Degree in Electronics and Communication Engineering from Madurai Kamaraj University and M Tech in Computer and Information Technology from Mononmanium Sundaranar University, Thirunelveli, Tamil Nadu, India. She is currently associated with TKM College of Engineering, Kerala, India as Assistant Professor and doing research in Indian sign Language Recognition System. Her areas of interest include Gesture Recognition, Signal and Image Processing and Embedded Systems. She is a member of IEEE and student councilor of IEEE Circuit and System Society, TKM College of Engineering, Kollam, Kerala. She is also a member of The Indian Society of Technical Education.

Dr. Shahul Hameed T A, received his Bachelor’s Degree in Electronics and Communication Engineering from TKM College of Engineering under University of Kerala, M. Tech in Micro-electronics and VLSI Design from IIT Kharagpur, India and Ph.D in Electronics and Communication Engineering from University of Kerala. He is having an experience of 23years in teaching and 8 years in industry. He is currently working as Professor in Electronics and Communication Engineering Department and Principal in TKM College of Engineering, Kollam, Kerala, India. He has authored more than fifty international journal and conference papers and has one patent in his credit. His areas of interests are Organic Devices, Quantum Devices and Device Modelling, and Mixed Signal Circuit Design.

Dr. O. Sheeba, was born in Thrissur, Kerala, India in October 20 1963. She received her B. Tech degree in Electronics and Communication Engineering from T.K.M College of Engineering, Kerala, India in 1987. She received her M. Tech degree from Cochin University of Science & Technology, Cochin, Kerala. She received her Ph,D degree from Kerala University, Kerala, India. She retired as Professor in the department of Electronics & Communication, T.K.M. College of Engineering, Kollam, Kerala, India. Dr. O. Sheeba is a member of The Indian Society of Technical Education and member of the Institution of Engineers, India