

# A Survey on Sign Language Detection

Shauvik Purkayastha  
Dept. of Computer  
Science & Engineering  
Assam Engineering  
College, Guwahati -  
781013, India

Mridul Jyoti Roy  
Assistant Professor  
Dept. of Computer  
Science & Engineering  
Assam Engineering  
College, Guwahati –  
781013, India

Navam Pradhan  
Dept. of Computer  
Science & Engineering,  
Assam Engineering  
College, Guwahati -  
781013, India

Prasurjya Sarma  
Dept. of Computer  
Science & Engineering  
Assam Engineering  
College, Guwahati -  
781013, India

## ABSTRACT

The most important aspect of a society is communication. Every culture in the world has their own way of communicating with people of their kind. But the people who cannot hear or speak cannot easily share their thoughts. This creates a sort of isolation for people like them. The hand gestures that people of special abilities make to communicate would be taken up from a camera and the same would be translated and show on the screen along with an audible source. The most basic methods one has to follow in most of the sign language detection system are tracking, edge detection, segmentation, and a dataset. The first and foremost part is the recognition of hand or motion tracking of the hand from the camera. After the hand is successfully detected the extra environment must be diminished. The removal of an extra environment is done with the help of edge detection and segmentation through various image processing algorithms. After the distinguished picture is obtained, with the help of machine learning techniques and neural networks the result obtained are compared to the existing normalized dataset which is then translated over and produced as an output.

## Keywords

Preprocessing, Edge detection, ANN, CNN, LDA, Skin detection, YCbCr

## 1. INTRODUCTION

People having physical limitations speech and hearing impairment is often unable to convey their message properly. A translator can be developed that will take hand gestures as input and give the equivalent alphabet as an output, which will enable those disabled people to communicate.

In order to determine the alphabet from the hand gesture than classification, noise reduction, pattern recognition and prediction for the hand gesture using real-time video need to achieve first. A standard sign language system needs to be chosen and apply an existing or a new dataset for it. The edge detection can be obtained with the help of various techniques like Canny, Sobel, Prewitt, Robert, Laplace, Frei Chen to name a few. And the segmentation part is obtained with the help of the YCbCr color model that helps in skin detection. The help of various machine learning technique and neural networks techniques like LDA (Linear Discriminant Analysis), SVM (Support Vector Machine), ANN (Artificial Neural Networks), CNN (Convolutional Neural Network) to name a few, are used for the training and processing of the obtained image. And after processing the output is either generated through a screen or a suitable audible source.

## 2. RELATED WORK

In [1], they have devised a method for the detection of the Indian sign language using hand tracking, motion tracking,

edge detection, and segmentation. The first module used there is the image acquisition module, where a video camera is used with standard specifications to acquire the images. The videos are then converted into frames, cropped in order to select a specific region and then an algorithm is implemented on it. The third module is hand tracking, where the difference between the 2 frames is determined and that part (the moving part) is retrieved and all the stationary parts are subtracted.

Along with that, the YCbCr skin color model is used in order to determine the part containing the skin. After the tracking is complete, hand segmentation is done, which is attained by techniques of thresholding and edge detection. After the extraction of the hand coordinates, template matching is done using recognition based on minimum distance to give the output.

In [2], Hybrid Wavelet Transform, fractional energy, Sign language recognition, Cosine Transform, and Haar Transform are used. Edge detection is achieved by various gradient operators like Roberts, Pretwitt, Sobel, Canny, Laplace, Frei-Chen and Kirsh operator. There are multiple orthogonal transforms available among which Cosine and Haar with fractional coefficients have shown the best performance for video retrieval. To evaluate the performance of classifiers, Euclidian distance, City block metric, cosine similarity, and Correlation are used.

Classifier Based Recognition detailed algorithm is:

1. Features of training sign images are extracted are given in the feature extraction phase.
2. Test feature set is evaluated on different classifier instances of KNN classifier created.
3. Based on the test image classification result, a sign is recognized from the test image set.
4. The accuracy of classier is calculated for each instance of the KNN classifier.

Skin detection, image preprocessing and different machine learning techniques like PCA and LDA and neural network for training and testing of the system is used [3]. Here they used approaches like Artificial Neural Network, fuzzy logic, genetic algorithms and others like PCA, Canonical Analysis etc. Skin filtering technique was used for the segmentation of the hand. Fang used adaptive boost algorithm which could not only detect single hand but also the overlapped hands.

In [4], webcam image acquisition and preprocessing feature extraction is used. The ANN method used here is a supervised learning method. The classification is done by using artificial intelligence. The processing is done by first converting the image into grayscale and then by thresholding. For skin

detection or skin segmentation, YCbCr algorithm is used. The skin color pixels are then used to detect the hand region. Edge detection is then achieved using Canny, Sobel and Prewitt. This system is implemented using MATLAB version 14.

In [5], Slope magnitude, Color Averaging, Row Mean, Column Mean, Forward Diagonal Mean, Tile Mean, and Triangle Mean are used. The dataset used here are the 26 alphabets of the English language. Finger spelling is done using hand and 26 hand gestures for each alphabet of the English language. For extracting edges, Sobel, Prewitt, Robert, Laplace, Canny, and Frei-Chen operators are used. The extraction process is divided into two parts: Feature extraction and Query Processing. For measuring the performance genuine acceptance ratio and false acceptance ratio is measured. This system is implemented in MATLAB version 12.0.

Object detection is achieved using Viola-Jones Algorithm, which is an algorithm that detects objects that are in the form of a hand [6]. The Viola-Jones algorithm detects the object using feature value which is faster than the detection based on the value per pixel of an image. Euclidean distance formula between two points is used to match the image with templates. The finger alphabets used there is the Standard Indonesian Sign Language which is often called SIBI. The system performs object detection to get the hand and then build a region of interest or area desired. After that, the recognition process is done. The conclusion of the tests gave the result that the detection process will be optimal if done at a distance between 30cm to 70cm from the lens and if the hands are held in upright position on 0 degrees.

Data is collected using leap motion controller, which is a USB peripheral device which is designed to allow users to control their computers using hand gestures alone [7]. This sensor can track hands, fingers, bones, and finger-like objects. The dataset taken was of a total of 520 samples consisting of 20 samples of each alphabet. The Dataset is normalized and then provided to the ANN for training and testing. This proposed system was able to detect most signs correctly and does not require a specific background or environmental conditions. The image recognition rate of this system was 96.15%.

Binarisation process [8] is used which converts an image into grayscale. The dataset used here is from the Korean alphabet. It does not use tracking, pattern recognition, classification etc. It uses preprocessed images with black background and hand part only. After binarisation, the center of gravity of the hand is calculated and then skeletonization is done, which converts the fingers and hand to lines. Using the skeletonized images, the template comparison is done to determine the alphabet.

ROI segmentation is used where a wide image is created at first by sampling and concatenating the original video frames [9]. And then by using the network that detects the hand area, the ROI segmented hand area is obtained. The second step is the sign language learning, where the ROI segment image is the input of the classification network, and the input data comes out as probability vectors. By using this, the sign is determined. The dataset used here is for 12 gestures in 60 situations taken from a 1m distance in various combinations of backgrounds, clothes etc. The success rate was calculated to be 84% without ROI and 97% with ROI segmentation in 1m tests.

In the 1.5m test set, the success rate was 55% and 83% respectively. Therefore, it was concluded that ROI

segmentation significantly increases the success rate and accelerates the training speed of the problem.

In [10], two types of gesture recognition systems are used. The first is touch based in which optical, magnetic or acoustic devices are attached to the hands which are used to report the position of the hands. This results in higher accuracy and latency. However, their demerit is that they require the user to wear some gear and carry cables connecting the device to a computer, making it not very viable for the regular users. The second is vision-based systems in which users have to stand in front of a camera and begin to deliver their message using gestures, which, even though has slightly lower accuracy, is a more viable method. The dataset used here were the Indian Sign Language dataset which contains about 7500 sign gestures, the DESIGN which is a Chinese language dataset, the RWTH German Finger spelling Database which is a German dataset containing 35 gestures, the MSR Gesture3D dataset which is a dynamic hand gesture dataset containing 12 dynamic hand gestures and a few others. Datasets were collected by Real-Sense and Kinect giving accuracies of 97.8% and 98.9% respectively.

In [11], the interpretation of sign language is divided into two phases, the first phase being the preprocessing phase, where the hand is extracted from the image by using background subtraction, filtering, noise removal, grayscale conversion, normalization, scaling and other such methods. The second stage is the classification stage, where the images are classified into the possible gestures using Haar Cascade Classifier. It is trained using 500 positives, 500 negatives and 50 test samples of each gesture. The next phase is the speech synthesis phase, where the gestures are converted into audio format. For that purpose, a text to speech converter is used. The input of the text to speech converter is the gesture that was recognized and the output is given in speech. The dataset used here was the Indian Sign Language Dataset. The accuracy obtained was 92.68%.

In [12], feature extraction and dimension reduction is done using principal component analysis. The system is training is done using the K Nearest Neighbor Algorithm and the Back Propagation Algorithm. The system was implemented in Matlab2017a and the testing was conducted with over 220 images of double handed Indian sign language alphabets and 800 images of single-handed Indian sign language alphabet. The accuracy achieved was 95.84% and 94.88% respectively.

In [13], firstly a video of a sign language demonstration is sampled and made into an image. After that, the image is given as the input of the convolutional neural network (CNN). This method uses only 2D images, hence it can be realized using a cheap camera and also the training data set is quite small as only sampled images are used. In the final test, the accuracy was found to be 86% using an inferior quality camera.

In [14], sign language gestures are captured by a depth camera. The process consists of 5 steps: preprocessing, projections, motion history images, and motion map extraction, 3D M-PHOG representation, and classification.

MHI (Motion History Image) is used for gesture representation to give the motion history by condensing the gesture sequence into a single grayscale image. The MHI is computed along the front, top and side views in order to increase the robustness of the system. The experiments are carried out on 2 data sets which are MSR Gesture 3d data set and the Design Dataset. The advantages of using the 3D-

MPHOG descriptor used here are that it can represent the sign language gestures in a more compact and discriminating way.

The approximation of arbitrary 2-D curves by polygon figures is performed [15]. It removes the unnecessary curves and bends and gives us a discrete figure that is bounded by line segments. Then convex decomposition is performed which gives us the minimum number of clusters ensuring that each cluster has a concavity lower than that of a predefined threshold. Thirdly, calculation of angle and distance with respect to wrist point is done. Here the aim is to find unique vectors to represent different hand gestures. Finally, generation and training of feature vector are performed. A feature vector of dimension 36 is created for each gesture, of which 36 partitions are made by dividing the scale from 0 degrees to 180 degrees in 5 equal divisions. In a given interval the lines joining the midpoint of a segment and wrist point is determined. The length of that line is stored for the particular nth interval. Finally, a matrix of features is created by combining the feature vector of all the gestures present in the training set and then the machine is trained with the matrix of features using SVM, artificial neural network, Naive Bayes and K-NN classifier and the accuracy is compared on a common test set.

In [16], end to end approaches that are used are convolutional neural networks (CNN), K-nearest neighbor (k-NN), decision trees, multi-layer perceptrons (MLP), support vector machine(SVM) and linear discriminant analysis (LDA). For the extraction process, PCA is used and different classifiers are used to do the recognition. The dataset used was the Irish Sign Language (ISL) dataset containing more than 50,000 images. Two approaches are compared here, the first of which is end to end and the second approach is based on feature extraction followed by different classifiers. The end to end classification approach includes the usage of CNN, LDA,

SVM, and MLP methods and the feature extraction based approach consist of using principal component analysis (PCA) to extract features which were then used as input to the classifiers k-NN, LDA, MLP, SVM, and Decision Trees. It was found in this paper that handcrafted features are competitive compared to features extracted using convolutional neural networks (CNN).

In [17], the classification method used is a multilayer neural network. 50 snapshots of each object of each object for neural network training set and 100 snapshots of each object for neural network testing set are used. In the first step, skin color pixels are detected, and then only the hand is extracted. Then segmentation is done using the obtained information. Binary images are created where white pixels represent the skin and black pixels are the rest of the image. Then the feature vector dimensions are reduced from 6-dimensions to 2D. For that purpose, the measure of scattering is used to reduce the dimensions of the image. Then the classification of signs is accomplished using a multilayer neural network. For training of neural network supervised learning is applied. The accuracy achieved in determining the signs in this method was 99.25%

### 3. METHODOLOGY

The process for sign language detection can be implemented efficiently by image processing or data set that is training set using methods like canny, Sobel, Prewitt etc, followed by machine learning methods such as ANN which is supervised learning method and LDA. For skin detection for the data set, the YCbCr algorithm is mostly preferred. Then a testing set will be used for the product testing. In fig 1, a flowchart of the methods is shown which is to be implemented for the detection of sign language and give text or audio output.

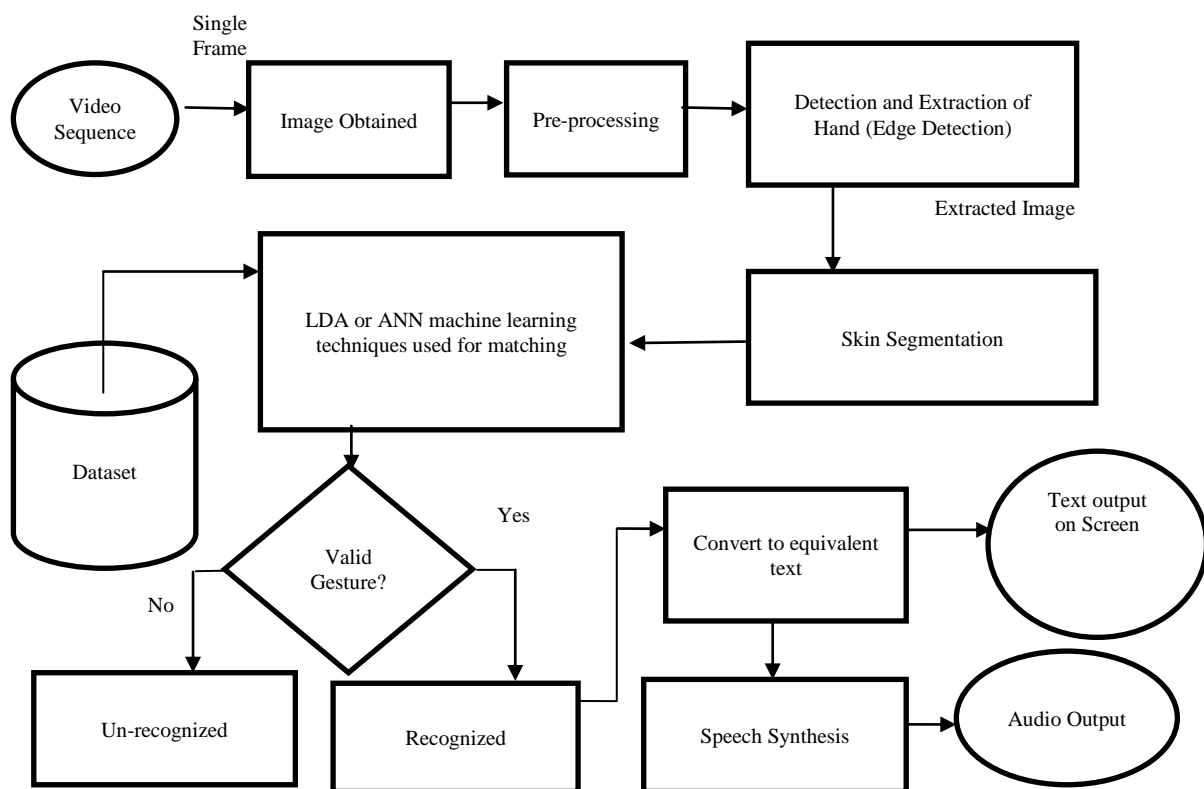


Fig 1: Flowchart

#### **4. CONCLUSION AND FUTURE WORK**

This paper displayed the different methodologies and strategies to detect the hand gesture and convert it to an audible or in text output. The Sign Language recognition system can be proposed where it can be further developed to detect advanced gestures for human-machine interface more efficiently. The recognition of sentences using hand gestures can be implemented in the future. Also, hand gestures combined with facial expressions can be used in the future to determine the emotions while detecting the words.

#### **5. REFERENCES**

- [1] Purva A. Nanivadekar and Dr. Vaishali Kulkarni, "Indian Sign Language Recognition: Database Creation, Hand Tracking and Segmentation", 2014 International Conference on Circuits, Systems, Communication and Information Technology Applicants (CSCITA).
- [2] Nalini Yadav, Sudeep Thepade and Pritam H Patil, "Novel Approach of Classification Based on Indian Sign Language Recognition using Transform Features", 2015 International Conference on Information Processing (ICIP) Vishwakarma Institute of Technology. Dec 16-19, 2015.
- [3] Rahat Yasir and Riasat Azim Khan, "Two-Handed Hand Gesture Recognition for Bangla Sign Language using LDA and ANN".
- [4] Priyanka C Pankajakshan and Thilagavathi B, "Sign Language Recognition System", IEEE Sponsored 2<sup>nd</sup> International Conference on Innovations in Information Embedded and Communication Systems ICIIECS'15.
- [5] Sudeep D. Thepade, Gandhali Kulkarni and Priti Kelvekar, "Sign Language Recognition using Color Means of Gradient Slope Magnitude Edge Images", 2013 International Conference on Intelligent Systems and Signal Processing (ISSP).
- [6] Setiawardhana, Rizky Yuniar Hakkun, Achmad Baharuddin, "Sign Language Learning based on Android For Deaf and Speech Impaired People", 2015 International Electronics Symposium (IES).
- [7] Deepali Naglot and Milind Kulkarni, "Real-Time Sign Language Recognition using the Leap Motion Controller".
- [8] Taehwan Kim and Sungho Kim, "Sign language translation system using latent feature values of sign language images", 2016 13th International Conference on Ubiquitous Robots and Ambient Intelligence (URAI) August 19-22, 2016 at Sofitel Xian on Renmin Square, Xian, China.
- [9] Sunmok Kim, Yangho Ji, and Ki-Baek Lee, "An effective sign language learning with object detection based ROI segmentation", 2016 13th International Conference on Ubiquitous Robots and Ambient Intelligence (URAI) August 19-22, 2016 at Sofitel Xian on Renmin Square, Xian, China.
- [10] Lihong Zheng, Bin Liang, and Ailian Jiang, "Recent Advances in Deep Learning for Sign Language Recognition".
- [11] Kanchan Dabre and Surekha Dholay, "Machine Learning Model for Sign Language Interpretation using Webcam Images", 2014 International Conference on Circuits, Systems, Communication and Information Technology Applications (CSCITA).
- [12] Kusumika Krori Dutta and Sunny Arokia Swamy Bellary., "Machine Learning Techniques for Indian Sign Language Recognition", International Conference on Current Trends in Computer, Electrical, Electronics and Communication (ICCTCEEC-2017).
- [13] Yangho Ji, Sunmok Kim, and Ki-Baek Lee, "Sign Language Learning System with Image Sampling and Convolutional Neural Network", 2017 First IEEE International Conference on Robotic Computing.
- [14] Lihong Zheng and Bin Liang, "Sign Language Recognition using Depth Images", 2016 14th International Conference on Control, Automation, Robotics & Vision Phuket, Thailand, 13-15th November 2016 (ICARCV 2016).
- [15] Himadri Nath Saha, Sayan Tapadar, Shinjini Ray, Suhrid Krishna Chatterjee and Sudipta Saha, "A Machine Learning Based Approach for hand Gesture Recognition using Distinctive Feature Extraction".
- [16] Marlon Oliveira, Houssein Chatbri, Suzanne Little, Noel E. O'Connor, and Alistair Sutherland, "A comparison between end-to-end approaches and feature extraction based approaches for Sign Language recognition".
- [17] Marko Z. Šušić, Saša Z. Maksimović, Sofija S. Spasojević and Željko M. Čurović, "Recognition and Classification of Deaf Signs using Neural Networks", 11<sup>th</sup> Symposium on Neural Network Applications in Electrical Engineering, NEUREL-2012, Faculty of Electrical Engineering, University of Belgrade, Serbia, September 20-22, 2012.