# Indian Sign Language Recognition Using ANN And SVM Classifiers

Miss. Juhi Ekbote
Final Year Student of ME (Computer Engineering),
B.V.M, VV Nagar, Gujarat, India

Mrs. Mahasweta Joshi
Assistant Professor
B.V.M, VV Nagar, Gujarat, India

***Abstract* – Sign language is an accepted language for communication between deaf and dumb community people. It is the most significant way of communication between normal people and hearing and speech impaired people without the need of an interpreter. Every country has its own developed Sign Language. In India, this dialect is known as Indian Sign Language. This research work aims at developing an automatic recognition system for Indian Sign Language numerals (0-9). The database used for implementation is self-created and consists of 1000 images, 100 images per numeral sign. Shape descriptors, Scale Invariant Feature Transform (SIFT) and Histogram of Oriented Gradients (HOG) techniques are used for extracting desired features. Artificial Neural Networks (ANN) and Support Vector Machine (SVM) classifiers are used to classify the signs. This system achieves accuracy as high as 99%.**

***Keywords:* Indian Sign Language, Shape descriptors, SIFT, HOG, SVM, ANN**

## I. INTRODUCTION

Sign Language is the dominant means of Non-Verbal Communication [12]. It is the way through which deaf and dumb community people can communicate with the outside world. Sign language uses gestures to communicate in place of voice. Sign language is a structured way of communication in which every word or alphabet is assigned to a particular gesture [2]. An assortment of communications via gestures exist over the world. The communication through signing utilized at a specific place relies on upon the way of life and talked dialect at that place [3]. A vast arrangement of effort has been done on British Sign Language (BSL), American Sign Language (ASL) and Chinese Sign Language Recognition [4]. Indian Sign Language (ISL) that contrasts extraordinarily from the previously mentioned SL measures has gotten minute thought so far.

Sign language is made up of a variety of gestures produced by diverse facial expressions and movements of hands or head/body. The gestures are chiefly divided into two classes: Static gestures and Dynamic Gestures. Static gestures comprise of only poses and configurations whereas dynamic gestures contain strokes, prestrokes, postures and phases. The dynamic gestures frequently include movement of body parts. It may also include emotions. There exists a variety of applications for Gesture Recognition System which are: Human Computer Interface, Video gaming, Augmented reality, Home appliances, Robotics, Sign Language, etc [1].

There are various methods for acquisition of gestural data. These methods are primarily divided into two types: Sensor based method and Vision based method. In sensor based method data glove or motion sensors are included from which the gestural data can be extracted [1]. Minute details are captured in this method. Moreover, this method requires wearing a hand glove with embedded sensors which makes the user little uncomfortable. In vision based method, the image is acquired with the help of camera(s). This method deals with the features of image such as color and texture that are mandatory for identifying the gesture. Indian communication through signing numerals are as shown in Fig. 1.

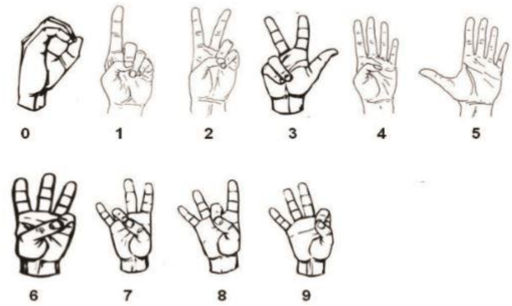


Fig. 1: Representation of ISL numerals (0-9) [3]

This paper involves five segments: segment II analyzes with reference to the past review made in the area of gesture recognition. The arranged methodology is shown in particular way in segment III. The exploratory outcomes are given in range IV. At last segment V gives the once-over and complete of the work.

## II. RELATED WORK

Purva C. Badhe and Vaishali Kulkarni proposed a system for recognition of ISL gestures from a video [1]. The proposed idea converts ISL numerals and alphabets into English. A combinational algorithm is adapted which includes Canny edge detection, YCbCr model for segmentation, thresholding, etc. for preprocessing step. Features are extracted using Fourier descriptors. The proposed scheme makes use of Euclidean distance for recognition of signs. The database used for implementation was self-created which included a total of 130,000 videos out of which 58,000 videos were tested for checking the performance of the system. An accuracy of 97.5% is achieved. The whole framework was produced in MATLAB by making a Graphical User Interface (GUI).

Madhuri Sharma, Ranjna Pal and Ashok Kumar Sahoo proposed a system for automatic recognition of ISL numerals [2]. Only a simple regular digital camera was used to acquire the gestures. A database containing 5000 signs, 500 images per numeral sign was created by them. Direct pixel value method and hierarchical centroid method were used to extract required features from images of numeric signs. Neural network and kNN classifiers were used to

978-1-5090-3294-5/17/$31.00 ©2017 IEEE

recognize the signs. The overall identification rate reported was 97.10%.

The work in [3], presents a method based on Artificial Neural Network capable of classifying static gestures of ISL numerals and alphabets automatically. The proposed method translates fingerspelling in ISL to textual form. The gestures included English alphabets (26 letters) and numerals (0-9). They used YCbCr model and also applied filtering and morphological operations for hand segmentation. A novel technique based on the distance transformation of the images was projected. Artificial neural network was used to categorize the gestures. The system was tested for 36 signs with 15 images of each. The system was implemented using MATLABR2010a. The proposed method had a low computational complexity. A recognition rate of 91.11% was achieved.

A. Singh, S. Arora, P. Shukla and A. Mittal in [4] proposed an idea which was capable of classifying ISL static gestures captured under indistinct conditions. Gestures were divided into single handed or double handed gestures. Geometric descriptors and HOG features were used to extract features. A database consisting of 260 images captured under simple and complex backgrounds was collected for the experimental purpose. The proposed approach compared KNN and SVM classifiers and it was concluded that SVM was superior to KNN algorithm in terms of accuracy on both geometric and HOG features. SVM achieved the highest accuracy of 94.23%.

The work in [5], presents a real time recognition of ISL gestures proposed by Shreyashi N. Sawant and M. S. Kumbhar. Otsu algorithm was used for segmentation purpose. PCA was utilized to diminish the element vector for a specific gesture picture. A dataset comprising of 260 pictures, 10 each of the 26 signs was utilized. The pictures were caught at a determination of 380 x 420 pixels. Euclidean separation was ascertained amongst test and prepare picture and signal having least separation was perceived. Perceived gesture was changed into content and voice configuration and content was shown on GUI screen.

K. Dixit and A. S. Jalal in [6] proposed an approach for translating ISL gestures into usual text. Global thresholding algorithm was used for segmentation purpose. Hu invariant moment and structural shape descriptors were used as feature extraction techniques. Multiclass Support Vector Machine was used as a classifier and maximum likelihood selection was done to recognize the gestures. A dataset consisting of 720 images was used. A recognition rate of 96% was achieved.

Ghotkar A. S., R. Khatal, S. Khupase, S. Asati and M. Hadap in [9] proposed their work which consisted of four modules for recognition of sign languages. CAM-SHIFT method was employed for hand tracking. HSV color model and neural network were used for hand segmentation. For feature extraction, Generic Fourier Descriptor (GFD) method was used. Genetic algorithm was used for gesture recognition. The authors did not mention about the database and no result was reported about their work.

## III. PROPOSED METHOD

This system includes following four major steps:

A) Data Acquisition
B) Pre-processing and Segmentation
C) Feature Extraction
D) Classification

### *A. Data Acquisition*

The database used in the implementation is self-created as the dataset of Indian sign language numerals is not available from any resources. Database is generated by capturing videos of Indian sign Language numerals (0-9). Videos are captured with the help of web camera integrated in the laptop. We used 1000 images, 100 images per numeral. The database consists of numerals from 0 to 9. The database consists of 10 signers. The database is divided into two sets, one set was used for training and other was used for testing.

### *B. Pre-processing and Segmentation*

After collecting the database from users, the images were pre-processed. This step is performed before extracting the features, so that the most useful information can be produced neglecting the redundant and noisy data. In order to recognize skin shading in the picture, it is at first changed over to YCbCr shading space [13]. YCbCr segregates RGB into luminance(Y) and chrominance sections (Cb and Cr) [3]. The function rgb2ycbcr is used to convert RGB images to YCbCr model which is available in Matlab environment. The result of segmentation produces a binary image. YCbCr is the most popular method for extracting skin portion of an image using chrominance values.

### *C. Feature Extraction*

Feature extraction is a form of dimensionality reduction [2]. In case of gestures, the input data i.e. in terms of frames, is too large to be processed and can have redundancies. That is why; input data has to be transformed into its reduced form [1]. After image preprocessing and segmentation, a binary image is obtained which contains the hand-shape signifying an exacting sign. In order to categorize this gesture, we need to take out definite characteristics of that image [3]. Many feature extraction techniques are available such as HOG [10], shape descriptors [10], SIFT [11], PCA [7], Fourier Descriptors [1], etc.

#### *1) Shape Descriptors*

The general state of the fragmented picture was separated via shape descriptors. It comprised of the accompanying components.

*a) Eccentricity [10]:* It is the proportion of the separation involving the foci of the circle and the extent of its real pivot. Its value is in the vicinity of 0 and 1. An oval whose value is 0 is really a circle, while an oval whose value is 1 indicates a line.

*b) Aspect Ratio [10]:* It indicates the proportion of significant pivot to small hub. It is valuable in isolating big and broad motions. It is computed as takes after:

Aspect ratio= (major axis)/ (minor axis) (1)

*c) Compactness [10]:* It depicts the picture closeness to its boundary regarding its middle. Its most extreme esteem is 1 for circles and for circular shapes its esteem diminishes. It can be computed as follows:

$$\text{Compactness}=(4.\wedge.\text{Area}_{\text{image}})/(\text{Perimeter}^2_{\text{image}}) \quad (2)$$

*d) Extent [10]:* It is the proportion of pixels in the locale to pixels in the aggregate bounding box. It can be characterized as follows:

$$\text{Extent}= (\text{Area}_{\text{image}}) / (\text{Area}_{\text{bounding box}}) \quad (3)$$

*e) Solidity [10]:* It determines the extent of the pixels in the curved body that are likewise in the district. It is defined as:

$$\text{Solidity}= (\text{Area}_{\text{image}})/ (\text{Area}_{\text{convexhull}}) \quad (4)$$

*f) Orientation [10]:* Scalar that indicate the edge (in degrees running from - 90 to 90) between the x-pivot and the significant hub of the circle that has an indistinguishable second-minutes from the area.

*g) Spreadness [10]:* It shows how the image is spread over the background. It can be computed as follows:

$$\text{Spreadness}= (\mu20*\mu02)/(\mu00*\mu00) \quad (5)$$

*h) Roundness [10]:* It is a measure of tallness and width of a turned insignificant jumping box. Its most extreme esteem is 1 for a hover and for different shapes its esteem is less. It is calculated as below:

$$\text{Roundness}= (4.\text{Area}_{\text{image}})/(\wedge.\text{major axis}^2) \quad (6)$$

*2) Scale Invariant Feature Transform (SIFT)*

It is an arrangement of descriptors characterized for a picture, which are brightening, turn and scale invariant [11]. SIFT is an algorithm used to detect and describe local features in images. This algorithm is useful in object recognition, image stitching, gesture recognition, etc. It is utilized to extricate particular invariant keypoints from images that can be utilized to perform dependable coordinating between various perspectives of a question or scene [10]. All Sign images have diverse number of keypoints [10].

SIFT can be utilized both as a component indicator and a descriptor [11]. To begin with, location, scale and orientation are computed for each keypoint. We have figured fundamental 25 keypoints for each sign gesture and scale and orientation values are put away as an element vector. From the experimental results, we found that SIFT does not provide a good enough accuracy. Only 24.60% accuracy is achieved when SIFT alone is used. But when it is combined with other algorithms, it performs better.

*3) Histogram of Oriented Gradients*

HOG is an ordinarily utilized descriptor for location of a question in a picture. This calculation makes utilization of nearby introductions and force appropriations for recognizing the state of a protest. This calculation checks the events of slope introduction in confined bits of a picture. The basic thought behind the histogram of arranged angles descriptor is that nearby protest appearance and shape inside a picture can be depicted by the dispersion of force slopes or edge bearings.

This algorithm is divided into 4 stages: Gradient computation, Orientation binning, Descriptor blocks and Block normalization. In this, the image is divided into small regions called cells and for pixels within each cell, a histogram of gradient directions is compiled. The segments are created over a piece and are harmonized [10]. To outline 81 features, 9 rectangular cells are used consisting of 9 bin histogram per cell.

The HOG descriptor has a couple key inclinations over various descriptors. Since it takes a shot at neighbourhood cells, it is invariant to geometric and photometric changes, except for dissent presentation.

*D. Classification*

Features extracted during the feature extraction phase are given as an input to classification step. A variety of classification techniques are available which are useful for recognizing the gestures. Classification step consists of two phases: training phase and testing phase. Image classification is the most important part of digital image analysis [2]. In our work, we have used Support Vector Machine and Artificial Neural Network for recognition of ISL numerals.

*1) Support Vector Machine (SVM)*

SVM is a regulated learning strategy. It was introduced by Vapnik [15]. In this technique, every information thing is spoken to as a point in a n-dimensional space with the estimation of each element being the estimation of a specific co-ordinate. The classification is then performed by finding the hyperplane that differentiates the two classes very well. SVM develops a hyper plane which has the biggest separation to the closest preparing information purposes of any class, since as a rule the larger the margin, the lower the generalization error of the classifier [15].

The objective of SVM is to create a model which predicts the class name of information cases in the testing set effectively. It utilizes bolster vectors (SVs) part to delineate information from info space to a high-dimensional element space which encourages the issue to be handled in straight shape [14]. SVM is computationally expensive and thus it runs slow. SVM produce very accurate results. It is robust to noise.

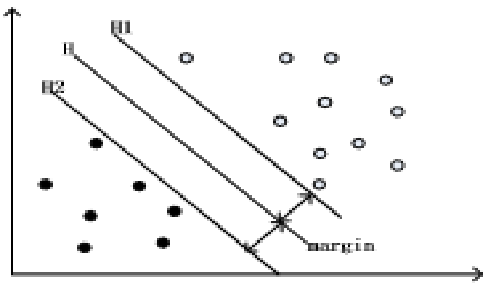


Fig. 2: Support Vector Machine [15]

*2) Artificial Neural Network (ANN)*

It is a biologically inspired model [3] stirred by the neural configuration of an individual personality. The taking care of parts in a produced neural framework are fake neurons which reflect the normal neurons. A manufactured neural system forms data by making associations between fake neurons. ANN is a well known machine learning calculation for performing acknowledgment [16]. The significant normal for neural system is its capacity to gain from past occasions or tests design [16]. In our proposed system, a feed forward neural network is used. The network consists of one input layer, one output layer and two hidden layers. The cases are shown to the framework through the neurons in the data layer and the yield of the structure is secured by the neurons in the yield layer. High accuracy is proficient by using ANN.

## IV. EXPERIMENTAL RESULTS

This segment provides the execution aftereffects of the gesture based communication acknowledgment framework. The framework was actualized utilizing MATLAB R2013a. Fig. 3(a), 3(b) and 3(c) show the snapshots of the proposed system.

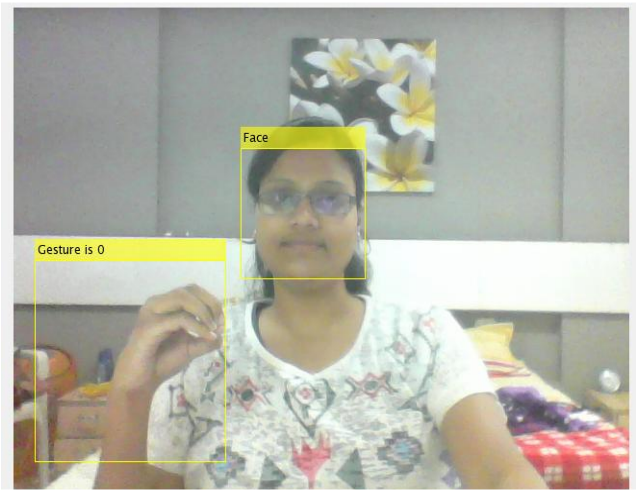


Fig. 3(a) shows the recognition of numeral 0

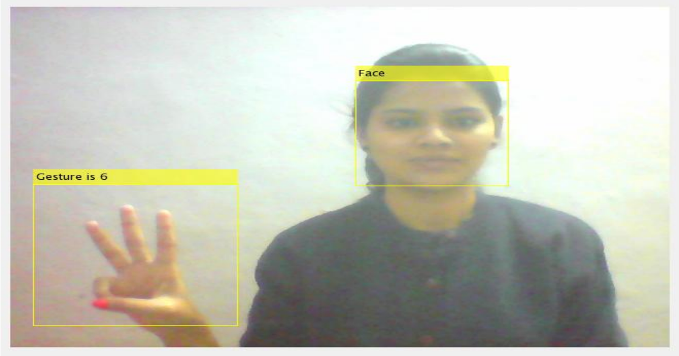


Fig. 3(b) shows the recognition of numeral 6

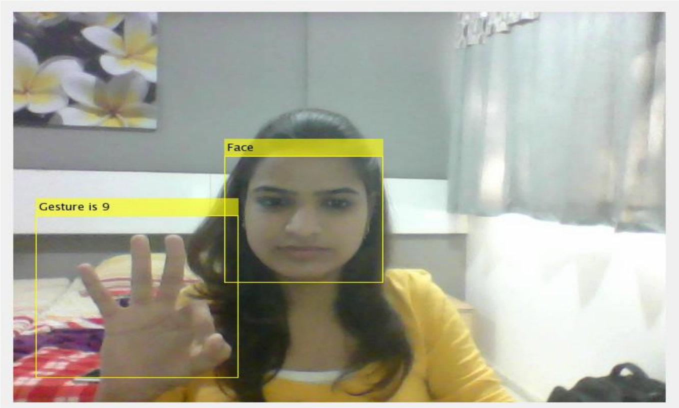


Fig. 3(c) shows the recognition of numeral 9

TABLE I demonstrates the acknowledgment precision of the framework.

TABLE I: RECOGNITION ACCURACY

| Experiment No | Feature Vector + Classifier | Accuracy (%) |
|---|---|---|
| 1. | Shape descriptors + SVM | 15.00 |
| 2. | SIFT + SVM | 24.60 |
| 3. | HOG + SVM | 96.20 |
| 4. | Combined | 93.00 |
| 5. | HOG + ANN | 99.00 |

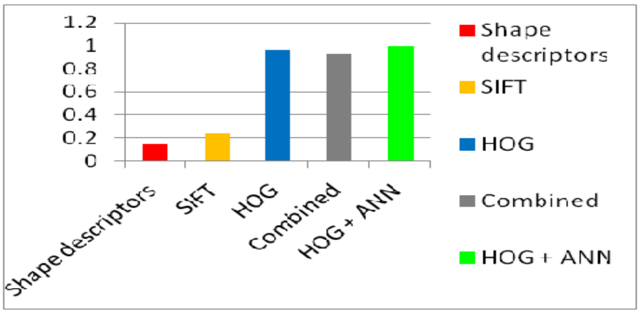


Fig. 4: Graph showing the accuracy obtained by applying different combination of feature extraction techniques and classifiers.

## V. CONCLUSION

In this paper, we have developed a recognition system for ISL numerals (0-9). The gestures were recognized using various feature extraction techniques like shape descriptors, SIFT and HOG individually along with SVM classifier and also we have combined all these features and found that this combined approach provides an accuracy of 93%. Experimental results demonstrate that a combination of HOG and ANN provides the accuracy as high as 99%.

## REFERENCES

[1] P. C. Badhe and V. Kulkarni, "Indian sign language translator using gesture recognition algorithm," 2015 IEEE International Conference on Computer Graphics, Vision and Information Security (CGVIS), Bhubaneswar, 2015, pp. 195-200.

[2] Madhuri Sharma, Ranjna Pal and Ashok Kumar Sahoo , “Indian Sign Language Recognition Using Neural Networks and KNN Classifiers”, ARPN Journal of Engineering and Applied Sciences, Vol. 9, No. 8, August 2014.

[3] V. Adithya, P. R. Vinod and U. Gopalakrishnan, "Artificial neural network based method for Indian sign language recognition," Information & Communication Technologies (ICT), 2013 IEEE Conference on, JeJu Island, 2013, pp. 1080-1085.

[4] A. Singh, S. Arora, P. Shukla and A. Mittal, "Indian Sign Language gesture classification as single or double handed gestures," 2015 Third International Conference on Image Information Processing (ICIIP), Waknaghat, 2015, pp. 378-381.

[5] Shreyashi Narayan Sawant and M. S. Kumbhar ,“Real Time Sign Language Recognition using PCA”, 2014 IEEE International Conference on Advanced Communication Control and Computing Technologies (lCACCCT),2014.

[6]K. Dixit and A. S. Jalal, "Automatic Indian Sign Language recognition system," Advance Computing Conference (IACC), 2013 IEEE 3rd International, Ghaziabad, 2013, pp. 883-887.

[7] Shreyashi Narayan Sawant. “Sign Language Recognition System to aid Deaf-dumb People Using PCA”, International Journal of Computer Science & Engineering Technology (IJCSET), Vol. 5 No. 05, May 2014.

[8] Geetha M, Manjusha U. C, "A Vision Based Recognition of Indian Sign Language Alphabets and Numerals Using B-Spline Approximation" International Journal on Computer Science and Engineering (IJCSE), March 2012.

[9] A. S. Ghotkar, R. Khatal, S. Khupase, S. Asati and M. Hadap, "Hand gesture recognition for Indian Sign Language," Computer Communication and Informatics (ICCCI), 2012 International Conference on, Coimbatore, 2012, pp. 1-4.

[10] S. C. Agrawal, A. S. Jalal and C. Bhatnagar, "Recognition of Indian Sign Language using feature fusion," Intelligent Human Computer Interaction (IHCI), 2012 4th International Conference on, Kharagpur, 2012, pp. 1-5.

[11] B. Gupta, P. Shukla and A. Mittal, "K-nearest correlated neighbor classification for Indian sign language gesture recognition using feature fusion," 2016 International Conference on Computer Communication and Informatics (ICCCI), Coimbatore, 2016, pp. 1-5.

[12] N. Baranwal, K. Tripathi and G. C. Nandi, "Possibility theory based continuous Indian Sign Language gesture recognition," TENCON 2015 - 2015 IEEE Region 10 Conference, Macao, 2015, pp. 1-5.

[13] Avinash Babu.D, Dipak Kumar Ghosh, Samit Ari; “Color Hand Gesture Segmentation for Images with Complex Background”; International Conference on Circuits, Power and Computing Technologies; IEEE 2013.

[14] L. Yun, Z. Peng, “An Automatic Hand Gesture Recognition System based on viola-Jones Method and SVMs”, International workshop on Computer Science and Engineering, IEEE Computer Society,2009,pp.72-76.

[15] L. Fei, W. Yajie, Q. Hongkun and W. Linlin, "Gender Identification Using SVM Based on Human Face Images," Virtual Reality and Visualization (ICVRV), 2014 International Conference on, Shenyang, 2014, pp. 444-448.

[16] O. A. Arigbabu, S. M. S. Ahmad, W. A. W. Adnan, S. Yussof, V. Iranmanesh and F. L. Malallah, "Gender recognition on real world faces based on shape representation and neural network," Computer and Information Sciences (ICCOINS), 2014 International Conference on, Kuala Lumpur, 2014, pp. 1-5.