Indian Sign Language Recognition System

Yogeshwar I. Rokade [#1], Prashant M. Jadav [#2]

[#] Department of Computer Engineering, Dharmsinh Desai University
Nadiad, Gujarat, India
[1] yogeshwarrokade55@gmail.com
[2] pmj.ce@ddu.ac.in

Abstract—**Normal humans can easily interact and communicate with one another, but the person with hearing and speaking disabilities face problems in communicating with other hearing people without a translator. The Sign Language is a barrier of communication for deaf and dumb people. People with hearing and speaking disability are highly dependent on non-verbal form of communication that involves hand gesture. This is the reason that the implementation of a system that recognize the sign language would have a significant benefit impact on dumb - deaf people. In this paper, a method is proposed for the automatic recognition of the finger spelling in the Indian sign language. Here, the sign in the form of gestures is given as an input to the system. Further various steps are performed on the input sign image. Firstly segmentation phase is performed based on the skin color so as to detect the shape of the sign. The detected region is then transformed into binary image. Later, the Euclidean distance transformation is applied on the obtained binary image. Row and column projection is applied on the distance transformed image. For feature extraction central moments along with HU's moments are used. For classification, neural network and SVM are used.**

Keyword - Artificial Neural Network, Central moments, Distance transformation, Fourier Descriptor, HU's moments, Indian sign language, Projection, Skin Segmentation, SVM.

I. INTRODUCTION

The sign language is used widely by people who are deaf-dumb; these are used as a medium for communication. A sign language is nothing but composed of various gestures formed by different shapes of hand, its movements, orientations as well as the facial expressions. These gestures are generally used by deaf-dumb people in order to express their thought. Dumb-deaf persons faces communication barrier in public places while interacting with normal person, such as in bank, hospital and post offices. Sometimes the deaf needs to seek the help of the sign language interpreter so as to translate their thoughts to normal people and vice versa. However, this way turns out to be very costly and does not work throughout the life period of a deaf person. So a system which can automatically recognize the sign language gestures becomes a necessity. Introducing such a system would lead to minimize the gap between deaf and normal people in the society. The sign language in use at a particular place depends on the culture and spoken language at that place. Indian sign language (ISL) is used by the deaf community in India. ISL is a standard and well-developed way of communication for hearing impaired people in India and speaking in English. Different symbols are involved for different alphabets for Indian Sign Language. It consists of both word level gestures and finger spelling. This paper presents a method for the automatic recognition of the static gestures in the Indian sign language alphabet. The signs considered for recognition include 17 letters of the English alphabet.

In the proposed approach, the main focus is on the classification and recognition of the Indian sign language given by the dumb-deaf user in real time. Thus, the speed and simplicity of the algorithm is important. The system approach involves segmenting the hand based on the skin colour statistics, then convert that segmented image into binary, apply feature extraction on the binary image, for extraction of the features the techniques used are distance transformation, Discrete Fourier Transform, Probability distribution property that is central moments [1], 7 moments [2]-[3] and for classification and recognition Artificial Neural Network [1] and SVM [4]-[5] are used. Fig. 1 shows the gestures used for all alphabets in the Indian sign language.

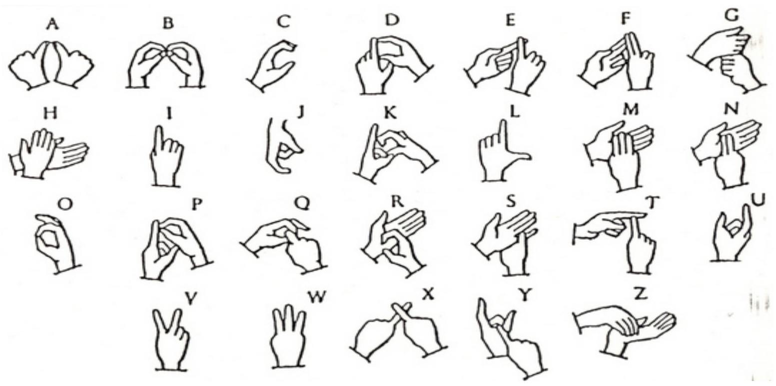


Fig. 1. Indian Sign Language

II. PAGE LAYOUT

In gesture recognition, the sign language recognition forms an important application. It consists of two different approaches [6].

- Glove based approach
- Vision based approach

1) Glove based approach: Here the signer requires wearing a sensor or a colored glove. Wearing the glove simplifies the task during the segmentation phase. The limitation to this approach is that it becomes mandatory for the signer to bear the sensor hardware including the glove during the entire operation.

2) Vision based approach: It makes use of the algorithms of image processing for detecting and tracking the hand signs including the signer's facial expressions. This vision based approach is simple since the signers need not wear additional hardware. In the proposed system vision based approach is used.

II. RELATED WORK

Adithya V, Vinod P. R, Usha Gopalakrishnan [1] presented in their work, Artificial Neural Network Based Method for Indian Sign Language Recognition. For segmentation RGB colour spaced are transformed into YCbCr color space, the pixel of skin colour in the input images are identified by applying a thresholding technique based on distribution of the skin colour in YCbCr colour space. The result of segmentation produces a binary image in which the skin pixels are white in colour and background in black colour. For feature extraction distance transformation, row and column projection applied on distance transformed image, Fourier descriptor is applied on row and column projected image. Central moments are calculated. Anchal Sood, Anju Mishra [7] have presented in their work, AAWAAZ: A Communication System for Deaf and Dumb. For segmentation they have used Hue-Saturation-Value (HSV) histogram. For the extraction of the features Harris algorithm is used. For Feature matching and recognition, the dataset already has the feature extracted of standard image and are stored as N*2 matrix mat file. The matrix value of this image query is then matched with each of those in the data set of every image and the minimum distance between the matched features is calculated to get the desired result. Shreyashi Narayan Sawant, M. S. Kumbhar [8] have presented in their work, Real Time Sign Language Recognition using PCA. Data acquisition: 260 images are used 10 images of each 26 signs. The algorithm used for segmentation purpose is Otsu's method. Noise is removed from the images using the morphological filtering techniques so as to get the contour. Here the main feature used is the principal component. In the phase of recognition, normalization is done for the subject gesture with respect to the average gesture and then it is projected onto the gesture space using the eigenvector matrix. At last, Euclidean distance is calculated between this projection and all the other known projections. The one being minimum value of these comparisons is chosen for recognition during the training phase. The recognized sign is converted to appropriate text and voice. Suriya M, Sathyapriya N,Srinithi M,Yesodha V [9] presented in their work, Survey on Real Time Sign Language Recognition System: An LDA Approach. The algorithm used for segmentation purpose is Otsu's method. Here the main feature used is the principal component. KNN classifier are used for classification and Similarity measures likes Euclidean distance, City Block Metric, Cosine Similarity and Correlation are made used so as to evaluate the performance of classifiers. Madhuri Sharma, Ranjna Pal and Ashok Kumar Sahoo [10] presented in their work, Indian Sign Language Recognition Using Neural Networks and KNN Classifiers. In their work first derivative Sobel edge detector method is used as it can compute gradient using the discrete difference between rows and columns of 3×3 neighbours. Feature extraction techniques used are direct pixel value and hierarchical centroid. For classification 2 classifiers are used that are: K-Nearest Neighbour (KNN), neural network pattern recognition tool.

III. PROPOSED APPROACH

A. Image Acquisition

Image acquisition is an operation of capturing the images of the hand gesture representing different signs. In this system publically available dataset is used for training and testing. The dataset used contains 17 different sign languages. The resolution for each image in dataset is 320*240. The resolution is same so as to lower the computational effort required for processing. The number of signs made use in the system are A, B, D, E, F, G, H, J, K, O, P, Q, S, T, X, Y, Z [13].

B. Hand Object Detection

1) Hand Segmentation: For skin detection adaptive probabilistic model is used. In this model manually annotated skin and background images are used for creating 32×32×32 RGB colour histograms for both skin and background appearance, and these histograms were normalized and used as probabilistic models of the skin and background [11]-[12]. First step is that the intensity level of RGB image is adjusted, later the adjusted input image is given as input to the skin detection model, the skin model further detect the skin region and convert the detected skin region into binary image. The skin value is determined and further normalized, and skin area is detected.

2) Filter and Noise Removal: The resulting binary image may include some sort of noise and error in segmentation. Filtering and various morphological operations are performed on the input image hence decreasing noise and errors in segmentation if any. Here image morphology algorithm is used that performs image erosion and dilation so as to eliminate the noise.

3) Feature Extraction: After image segmentation and pre-processing, binary image is obtained containing the shape of hand which represents a particular sign. In order to classify this image, extraction of certain features of that image is employed. Shape is considered to be an important visual feature of an object. In this work feature for shape representation is used. The proposed shape feature is derived from the distance transform of binary image [1].

- Distance Transformation: It is a derived representation of an image which is normally applied upon binary images [1]. To apply distance transform on an image, it should be first converted to binary form. A binary image contains object pixels as well as non object pixels. Applying distance transform of such image gives another image of the same size where each pixel value is replaced by the minimum distance of that pixel from its nearest background pixel. So it results in a gray-scale image where the gray scale intensity of the foreground region corresponds to the distance from the closest boundary pixel. In the proposed work, the distance transform is computed by using the Euclidean distance.

$$d_e(\text{P, Q})=\sqrt{(x-u)^2+(y-v)^2}, \text{ where P and Q are two points} \quad (1)$$

- Projections of Distance Transform Coefficients: This step calculates the row projection vector and the column projection vector from the distance transformed image [1]. The calculation of projection vectors works as follows:

-The summation of the pixels values in the row and column of the distance transformed image is calculated. This step returns two vectors.

-The first value is the Row vector, say R, where each value in this vector is the total sum of non- zero pixel values of the row from the distance transformed image.

-The second value is the Column vector, say C, where every element in the vector is the total sum of non-zero pixel values of the column of the distance transformed image.

-The above steps give the 1-D row projection and column projection vectors, which uniquely represent the shape of the hand from the input image. These vectors are considered to be the shape descriptor. These shape descriptors represent the shape of the hand locally, but are sensitive to noise. So these descriptors have to be processed further to make them robust.

- Fourier Descriptors: Fourier descriptors of the shape are formed by the Fourier transform coefficients of the shape descriptors [1]. Here in this work Fourier descriptors for the row and column projection vectors are computed. For the two vectors R (t) and C (t), where t=0, 1, 2... N-1 the Discrete Fourier Transform is given by:

$$v_n=\frac{1}{N}\sum_{t=0}^{N-1} R(t)exp\left(\frac{-j2\prod nt}{N}\right), \text{ where n=0, 1, 2…N-1} \quad (2)$$

where N is size of R.

$$u_n=\frac{1}{N}\sum_{t=0}^{N-1} C(t)exp\left(\frac{-j2\prod nt}{N}\right), \text{ where n=0, 1, 2 …N-1} \quad (3)$$

where N is size of C.

The coefficients u_n and v_n are the Fourier descriptors of the shape.

- Feature Vector: The values of feature are generated from the Fourier descriptors of the row and column projection vectors by considering only the magnitude value of the Fourier coefficients and neglecting the phase information. The normalization of the feature values is done by dividing the values of magnitude for every Fourier coefficients by the value of magnitude of only the first coefficient known as the dc component. This feature vector for every gesture comprises of 6 feature values which are the 2^{nd}, 3^{rd} and 4^{th} central moments of normalized Fourier coefficients of projection vectors of row and column [1].
- Central Moments: They are a set of values which characterize the properties of the probability distribution. The central moments of higher order are related only to the shape and spread of the probability distribution not to its location. For any real-valued random variable, say X, the k^{th} moment about the mean or k^{th} central moment is given by μk= $E\lfloor (XE[X])^k \rfloor$ where E denotes Expectation operation. The zeroth central moment μ0 is one. The 1^{st} central moment μ1 is zero. The 2^{nd} central moment μ2 is known as the variance usually denoted as σ^2, where σ denotes the standard deviation of the distribution. The 3^{rd} central moment μ3 defines skewness and 4^{th} central moment μ4 defines kurtosis.
- Hu moments: Hu invariants moments [2] are calculated by using geometrical moments of hand region. The first 6 Hu moments gives shapes which are invariance to translation, scale and rotation. The 7^{th} Hu moment gives shape which is skew invariance, and helps to distinguish between mirrored images.

4) *Classification:* The feature vector obtained from the feature extraction step is used as the input of the classifier that recognizes the sign. Artificial neural network is used as the classification tool. Classification step involves two phases: training phase and testing phase.

- Artificial Neural Network: A feed forward neural network in combination with a supervised learning scenario is used in the proposed method.
- Training Phase: The neural network is trained to classify 17 hand signs. The training dataset comprises 548 images including all the 17 signs.
- Testing Phase: In this phase, a dataset containing 300 images are used for testing which includes all the 17 signs.
- SVM: Here polynomial kernel is used. It is because after attempting several executions it was observed that polynomial kernel seeks to provide better result when compared to other kernel.
- Mathematically, Polynomial kernel

$$\mathrm{K}(x_{i,}x_j)=\left(x_i^T x_j + 1\right)^q, \text{where q is the degree of the polynomial} \quad (4)$$

-Training Phase: A total of 479 training data is used.

-Testing Phase: A total of 369 testing data is used.

IV. IMPLEMENTATION DETAILS

A. Implementation Platform Details

The hardware and software specifications of the platform on which the proposed approach implemented and tested is given below:

1) *Hardware Specification*
   - Processor : Intel(R) Core(TM) i3-2370M CPU @ 1.80GHz
   - RAM : 4.00 GB RAM
   - System :64-bit Operating System
2) *Software Specification*
   - OS :Windows 8
   - Front End: MATLAB R2012a

B. Dataset Details

The dataset are acquired from the internet with all the images having black background. The input image consists of only sign language gesture and no other skin area is present. The total number of images used is 848 with 320*240 of dimension. Below Fig. 2 shows some sample images from the dataset. Among 848 images, 548 images are used for training and 300 images are used for testing.

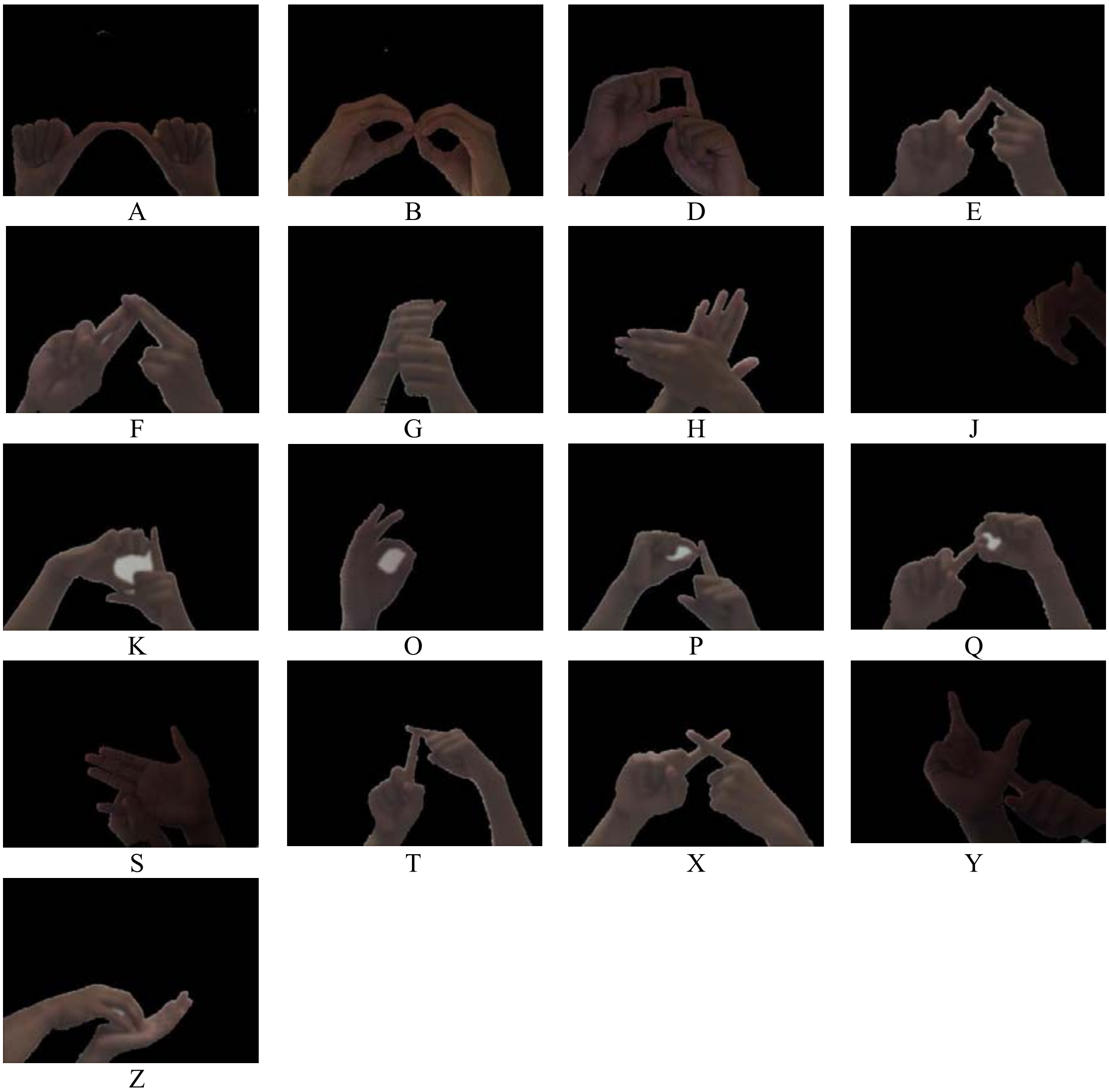


Fig. 2 Sample Dataset Images of Different ISL

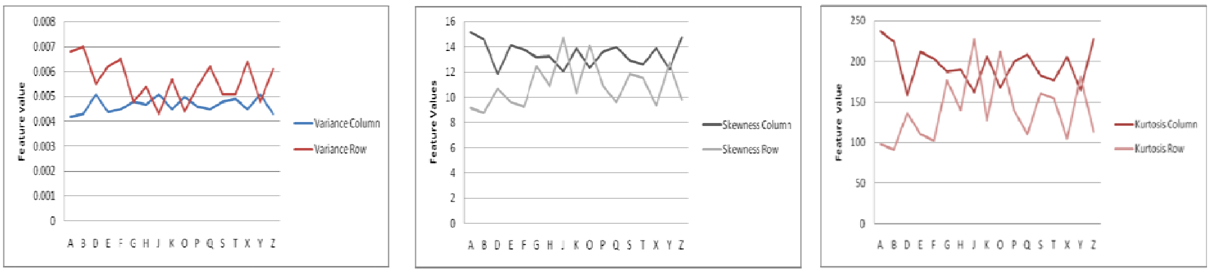
C. Experiment Results

13 features were extracted from each sign language which used as feature vector. Table I below shows the 6 feature values of 17 different gestures.

TABLE I. Six Central moments for 17 Signs

Sign	VarC1	VarR1	SkwC1	SkwR1	KurC1	KurR1
A	0.0042	0.0068	15.189	9.1397	236.4197	98.1084
B	0.0043	0.0070	14.6303	8.7971	223.8076	91.8308
D	0.0051	0.0055	11.8636	10.7263	158.0444	136.3457
E	0.0044	0.0062	14.1241	9.6191	210.9825	110.6914
F	0.0045	0.0065	13.7734	9.2588	202.8136	102.778
G	0.0048	0.0048	13.2152	12.4637	187.1120	176.4932
H	0.0047	0.0054	13.2800	10.9552	189.6922	140.4777
J	0.0051	0.0043	12.0618	14.7778	161.7575	227.1586
K	0.0045	0.0057	13.8885	10.3828	206.1105	127.9167
O	0.0050	0.0044	12.4051	14.0806	167.5301	211.5538
P	0.0046	0.0054	13.6152	10.9229	199.0306	140.2061
Q	0.0045	0.0062	13.9727	9.6071	207.9608	110.2707
S	0.0048	0.0051	12.9601	11.8738	182.0851	160.8284
T	0.0049	0.0051	12.6671	11.5687	176.2260	154.7142
X	0.0045	0.0064	13.8649	9.3682	205.1218	105.4049
Y	0.0051	0.0048	12.2291	12.7820	164.2606	181.3414
Z	0.0043	0.0061	14.7880	9.8542	227.4055	114.0058

The figure below shows comparison between row and column projection of Variance, Skewness and Kurtosis. Fig 2 (a) shows comparison between row and column projection of Variance of 17 different signs, Fig 2 (b) shows comparison between row and column projection of Skewness of different 17 signs Fig 2 (c) shows comparison between row and column projection of Kurtosis of different 17 signs.



(a) (b) (c)

Fig. 2. Comparison of row and column projection for different central moments

The feature vector contains 13 features, the first 6 features are Central Moments and the remaining 7 are Hu's moments. Table II below shows 7 Hu's moments of 17 different signs.

TABLE II. Seven HU's Moments for 17 Signs

Sign	M0	M1	M2	M3	M4	M5	M6
A	0.6381	0.2916	0.0056	9.1253	-3.6031	-3.5826	-2.1693
B	0.4109	0.0866	0.0145	0.0016	2.7850	-4.5987	3.2022
D	0.3106	0.0208	0.0121	0.0011	-2.4237	-1.2189	-7.0560
E	0.3616	0.0489	0.0071	0.0022	1.9818	1.4252	1.3493
F	0.3531	0.0435	0.0122	0.0015	-4.4885	-1.4169	-5.3720
G	0.2847	0.0041	0.0133	5.6207	-4.3144	9.9616	1.1319
H	0.2140	4.491	0.041	1.9381	7.0079	-3.6121	-2.5584
J	0.7519	0.0206	0.0736	0.0620	3.2216	-0.0077	0.0012
K	0.4400	0.0711	0.0352	0.0046	-1.4325	-4.0742	5.5091
O	0.2290	0.0107	3.9331	1.5908	-1.5122	-5.5950	3.7250
P	0.5732	0.0870	0.1385	0.0186	-7.1204	-0.0026	-8.7929
Q	0.4856	0.0842	0.0371	0.0031	-2.2606	-8.8984	-1.2767
S	0.2047	0.0023	9.9797	5.1654	4.0561e	-1.4779	-7.5074
T	0.4553	0.0235	0.0444	0.0045	-2.8120	-5.4845	-4.9772
X	0.4085	0.0689	0.0263	0.0028	-2.1128	-5.2509	-2.0538
Y	0.3372	0.0203	0.0294	0.0127	1.4071e	0.0015	1.9126
Z	0.3834	0.0891	0.0014	5.1961	-1.3978	4.2731	2.6204

Various Experiments Performed:

- Classification using ANN
  - -Using 6 feature sets
  - -Using 8 feature sets
  - -Using 13 feature sets
- Comparison of Accuracy result of different feature set in ANN
- Classification using SVM
  - -Using 6 feature sets
  - -Using 8 feature sets
  - -Using 13 feature sets
- Comparison of Accuracy result of different feature set in SVM
- Comparison of ANN Accuracy result with SVM Accuracy result

Table III below shows the Result of ANN.

TABLE III. Average Accuracy for Different Feature Set Using ANN

Features	Average Accuracy
6 features	88.3%
8 features	91.94%
13 features	94.37%

Dataset is reshuffled 10 times and are used for testing and training, and later average accuracy is considered. Below Fig. 3 shows the average accuracy using different feature set in ANN.

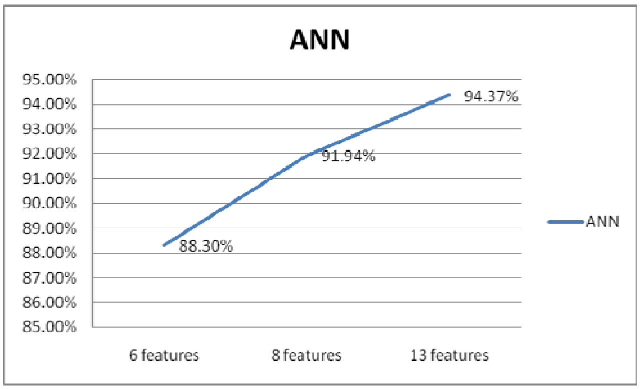


Fig. 3 Graph Showing Average Accuracy Comparison using different Feature Set in ANN

Table IV below shows the Result of ANN. The average accuracy is 92.12%. The result is improved from 69.92% accuracy with using 6 features and 89.33% accuracy with using 8 features, to 92.12% accuracy using 13 features.

TABLE IV. Average Accuracy for Different Feature Set Using ANN

Features	Average Accuracy
6 features	69.92%
8 features	89.33%
13 features	92.12%

The dataset is reshuffled 10 times and are used for testing and training, and later average accuracy is considered. Below Fig. 4 shows the average accuracy using different feature set in SVM.

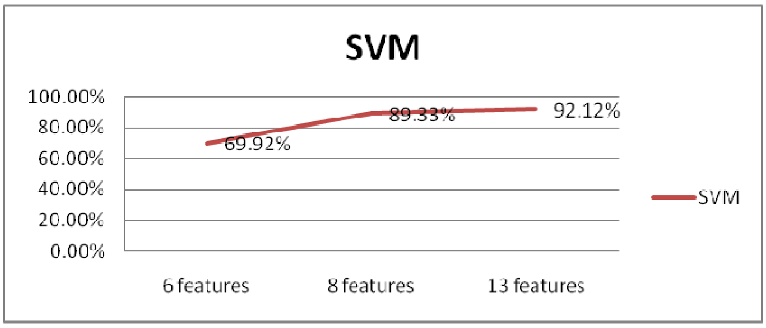


Fig. 4 Graph Showing Accuracy Comparison using different Feature Set in SVM

Table V below shows Comparison between ANN and SVM results.

TABLE V. Accuracy Comparison between ANN and SVM

Feature set	ANN Average Accuracy	SVM Average Accuracy
6 features	88.3%	69.92%
8 features	91.94%	89.33%
13 features	94.37%	92.12%

Below Fig. 5 shows the Result of Comparison between ANN and SVM

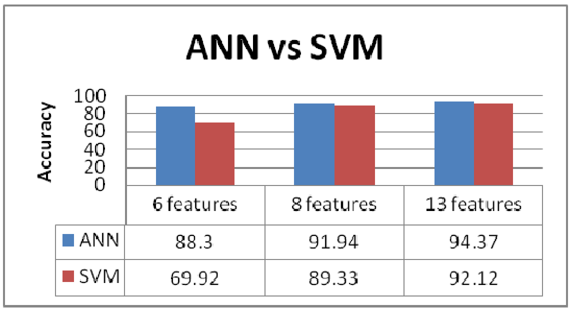


Fig. 5 Graph showing Accuracy comparison between ANN and SVM

By this experiment it is observed that ANN (Artificial Neural Network) gives higher accuracy of 94.37% with 13 features over SVM with 92.12%.

V. Conclusion

A novel approach to recognize the Indian sign language using Artificial Neural Network (ANN) and Support Vector Machine (SVM) is presented in the paper. The sign language used by deaf-dumb people are not understood by normal people and face a lot of difficulty, the proposed system can be used to understand the meaning of Indian sign language. Skin segmentation is used to get shape of the hand region, Euclidean distance transformation is employed so as to create grey level image and feature vectors are created using feature extraction. For feature extraction Central moments and HU moments are used. Artificial neural network is used to classify the sign which gives average accuracy of 94.37% and SVM classifier for the same gives accuracy of 92.12%. Both the classifiers give higher accuracy with 13 features. ANN gives better accuracy even with less number of feature set.

References

[1] Adithya, V., P. R. Vinod, and Usha Gopalakrishnan. "Artificial neural network based method for Indian sign language recognition" IEEE Conference on Information & Communication Technologies (ICT), IEEE, 2013.
[2] Conseil, Simon, Salah Bourennane, and Lionel Martin, "Comparison of Fourier descriptors and Hu moments for hand posture recognition" Signal Processing Conference, 15th European. IEEE, 2007.
[3] Sabhara, Reza Kasyauqi, Chin-Poo Lee, and Kian-Ming Lim, "Comparative Study of Hu Moments and Zernike Moments in Object Recognition" SmartCR 3.3, 166-173, 2013.
[4] Kurdyumov, Ruslan, Phillip Ho and Justin Ng, "Sign Language Classification Using Webcam Images", 2011.
[5] Singh, Akanksha "Indian Sign Language gesture classification as single or double handed gestures", Third International Conference on Image Information Processing (ICIIP), IEEE, 2015.
[6] Pandey, Pratibha, and Vinay Jain "Hand Gesture Recognition for Sign Language Recognition: A Review" Hand 4.3, 2015.
[7] Sood, Anchal, and Anju Mishra "AAWAAZ: A communication system for deaf and dumb" Reliability, 5th International Conference on Infocom Technologies and Optimization (Trends and Future Directions) (ICRITO), IEEE, 2016.
[8] Sawant, Shreyashi Narayan and M. S. Kumbhar, "Real Time Sign Language Recognition Using PCA", International Conference on Advanced Communication Control and Computing Technologies (ICACCCT), IEEE, 2014.
[9] Rajam, P. Subha, and G. Balakrishnan. "Real Time Indian Sign Language Recognition System to Aid Deaf-Dumb People" IEEE 13th International Conference on Communication Technology (ICCT), IEEE, 2011.
[10] Sharma, Madhuri, Ranjna Pal and Ashok Kumar Sahoo, "Indian Sign Language Recognition Using Neural Networks and KNN Classifiers" ARPN Journal of Engineering and Applied Sciences 8, 2014.
[11] Conaire, Ciarán O., Noel E. O'Connor, and Alan F. Smeaton. "Detector Adaptation by Maximising Agreement between Independent Data Sources" IEEE Conference on Computer Vision and Pattern Recognition, 2007. CVPR'07, IEEE, 2007.

AUTHOR PROFILE

Yogeshwar I. Rokade has completed Master of Technology in Computer Engineering from Dharmsinh Desai University, Nadiad, India. His main area of interest is in image and video processing.

Prof. Prashant M. Jadav is working as an Associate Professor in Computer Engineering department at Dharmsinh Desai University, Nadiad, India.