Recognition of Alphabets of Indian Sign Language by Sugeno type Fuzzy Neural Network

Shweta Dour[a], Dr . M.M.Sharma[b]

[a]Research Scholar, Bhagwant University , Ajmer, 305001, India
[b]Principal, Govt. Engineering College, Ajmer, 305001, India

Abstract- Sign Language Recognition has evolved as an important area of research in the past few years. Sign language can be defined as the language of the deaf and dumb people by which they are able to express their thoughts. Such people are not able to use acoustic means for the purpose of communicating, instead they convey message by making use of the Sign Language. Thus Sign language is a means developed for the deaf and dumb society by which they can visually transmit different sign patterns to convey their message by combining simultaneously hand shapes, movement of hands and orientation of hands which are also sometimes associated with the facial expressions. Like all other languages, Sign Languages also have their own alphabets and grammar. As these sign languages are not much known to the people outside the deaf and dumb communities and thus there always exists a gap between the impaired people and the normal people who have the abilities to talk and listen. The techniques developed for recognizing signs would allow the creation of systems which can help to eliminate this barrier, by providing computer tools to assist in the learning of sign language and in creating a system to translate the sign language to text for understanding of normal people. This would help a hearing impaired person to easily interact with a normal person at different levels in the society. Sign Languages are not same throughout the world. These are different for different parts of the world. The Sign Language used in India is called as the Indian Sign Language. In this paper we attempt to summarize the significance of features associated with the alphabets of the Indian Sign Language.

Index Terms- Indian Sign language (ISL), Skin Locus, Adaptive Neuro Fuzzy Inference System, Sugeno Fuzzy Inference System.

I. Introduction

A number of people in our society across the world experience difficulties in terms of communication such as being totally deaf and dumb or only dumb or sometimes only deaf and thus are not able to make the normal people understand their thoughts. The only way they communicate is through specific gestures which have particular meaning and these meaningful gestures are called as Sign Languages.

Sign language is considered as the basic means of communication among the deaf and dumb people. According to the International Journal of Language and Communication Disorders “Sign language can be considered as a combination of postures, facial expressions movements, and gestures equivalent to words and letters for natural languages.” In other words we can say that Sign languages are complete natural languages, with their own syntax and grammar.

But these languages are not much popular besides the Deaf and Dumb communities, and due to this reason there always exists a barrier between the people of hearing impaired communities and hearing people.

In general, to bridge the communication gap between the deaf and the hearing people, a lot of effort has been put across the world for different sign languages to produce a translation system that translates sign into spoken language and vice versa.

Developing the methods for recognizing signs would be useful to eliminate the communication barrier among the two communities. These methods would include the computer aided systems to provide sign-language to text translation systems or vice versa.

There is no internationally recognized and standardized sign language for all deaf people. As is the case in spoken languages, every country has got its own sign language with high degree of grammatical variations. The well known sign language used in India is commonly known as Indian Sign Language (ISL) [1]. Indian Sign Language (ISL) developed and used by the deaf and dumb communities in India not only acts as a mode of communication among deaf people but also it is a requirement of the other hearing people related to such people to know and understand their language for example parents of hearing impaired child would necessarily require to learn the sign language. Indian Sign Language comprises of static gestures that is for alphabets and for some ‘names of animals’, for word ‘stop’ for word ‘this’ etc. and also dynamic gestures that are gestures involving movement of hands. Thus to recognize them it is required to find the features for different static and dynamic states of hands which may include either movement of both hands, different shapes of hands in contact to other body parts or sometimes one hand moving faster as compared to another hand.

Towards learning any language the learning of alphabets of the language is the first and foremost step, similarly towards recognizing the Indian Sign language we have considered recognition of alphabets as the primary requirement such that we are able to develop a system to be able to recognize the alphabets and convert them text. Thus building a system to recognize signs of the Indian sign language would definitely be useful to the society. Therefore in this paper we would summarize about the recognition of alphabets i.e. A to Z and for Indian Sign Language, wherein we will be acquiring the input from an i-ball C8.0 web camera.

II. Related work for other Sign Languages

A number of researchers have proposed various different methods for sign language recognition systems. Usually these type of systems are developed by two approaches that is the Glove based approaches and the other is the vision based approaches. The glove based approaches do not require the detection of the hand as the hands are associated with the sensor gloves whereas the vision based analysis are based on the inputs images captured by cameras. Vision based analysis are comparatively user friendly as the users need not wear any type of bulky gloves over the hands and thus are more natural for developing sign language recognition systems. Though being user friendly the vision based systems are always associated with the complexities and challenges associated to the detection of bare hands from the camera inputs. Such vision based systems would initially require the detection and segmentation of hands from the acquired images. As we have concentrated over developing a vision based system so we provide the related work for the vision based analysis in this section.

Agris[2] proposed an isolated sign recognition system based on a combination of Maximum A Posteriori estimation and Maximum Likelihood Linear Regression. They developed a method to specifically recognize the single or one handed signs. They implemented selected adaptation methods from speech recognition to improve the performance of their system when performing user independent recognition. A recognition rate of 78.6% was reported when recognizing 153 isolated signs.

Juang[3] proposed a method for gesture recognition that was based on fuzzy temporal sequencing by the use of Fuzzified Takagi-Sugeno-Kang (TSK)-type Neural Network for hand gesture recognition. They applied their method to a gesture recognition task and experiments showed a recognition rate of 92%.

Fang[4] addressed the problem of large vocabulary sign recognition by proposing a combination of self organizing feature maps, HMM's hierarchical decision tree, with low computational costs, for the recognition of isolated signs. Experiments were conducted on a data set of 61365 isolated samples of 5113 different signs. Results showed an average recognition rate of 91.6%.

Byung[5] , presented the recognition of static gesture obtained from the images considering a 2D image plane. No other sensors or devices were considered associated with the hands. The recognition of static gestures was accomplished by the use of image moments obtained from the hand posture, and for dynamic gestures the movement trajectories for hands were considered and were recognized by use of Hidden Markov Models (HMMs) .

Another example of a trainable hand-gesture recogniser was the work by Murakami &Taguchi [6] who used neural networks to recognise the Japanese kana manual alphabet because of the network's learning abilities. They attempted the work in two parts: the first part was to recognise postures (static hand positions) and the second part is to recognise gestures(moving hand and finger motions), which they claim were more difficult to recognise because of the motion. For posture recognition, Murakami and Taguchi used a backpropagation neural network with 13 input nodes, a hidden layer of 100 nodes and an output layer of 42 nodes. Each of the 13 input nodes represents one of the data items that the VPL DataGlove measured (10 finger angles, x, y, z, and yaw, pitch and roll). The 42 output nodes represent the 42 postures of the manual alphabet that were to be recognised. They trained their network with 206 learning patterns and it achieved a recognition rate of 98.0% for the registered trainer, and a recognition rate of 77.0% for gestures made by people that the network had not been trained on.

T. Shanableh[7] developed a method for the recognition of isolated Arabic sign language. The system developed required the users to wear gloves so that the segmentation of hands becomes easier by the use of color segmentation. The system was user independent and the features extracted by the system were utilized for classification by two different approaches that were the K-NN and polynomial networks.

III. Related work for Indian Sign Languages

Attention towards the linguistic studies of ISL was started from the year 1978 onwards. Since then the ISL got the acceptance nationwide that it is the language of the deaf and dumb community of India having its own syntax , grammar and phonetics.

Years after its acceptance as the sign language of the nation in the later decade of 2000, research began on ISL recognition system. Also there has not been found any remarkable work done in the area of Indian Sign language recognition systems. A few have been summarized below.

Geetha[8] proposed a system for recognition of ISL by the use of piecewise polynomial functions to approximate the contours/surfaces by the use of small number of control points. The use of B-Spline approximation was adopted for the purpose of shape matching of ISL alphabets & numerals. Boundry tracing of the sign gesture was achieved in order to obtain the control points which were the Maximum curvature points. Next the control points were fitted with the piecewise continuous parametric polynomial functions called B-Spline curves. The thus obtained B-Spline curve for the purpose of sampling and smoothening undergone a set of iterations. After a series of iterations the maximum curvature points contributed to the shape of the gesture. The 2D space of the gesture was divided into 8 octants. Each octant was counted for the number of key points and thus the feature vector containing of set of 8 values each corresponding to the count of Maximum Curvature Points. Lastly the SVM classifier was utilized for recognition. The designed system achieved recognition rate of about 80% on an average for the static alphabets and numerals.

P.Subha Rajam[9] presented a system for the the 32 combinations of binary number sign by the use of specifically right hand palm. The system was designed to scan the input image of right hand palm in order to identify the finger tip positions named as little fingers, ring finger, middle, index finger and thumb finger. To identify the finger they used the measurement of heights of the finger tips with respect to a reference point considered at the bottom of the palm close to the wrist. Euclidean distance measurement was used for the determination of these heights from the reference points. The points that were extracted by the edge images were assigned the binary code based on the 'UP' or 'DOWN' positions of the fingers. The overall average recognition rate for two datasets was about 70- 80 %.

Research on Indian Sign Language made by Pravin Futane and Dr.Rajiv[10] used feature extraction based on shape and geometry feature and lastly learning by General Purpose Fuzzy MinMax(GFMM)neural network.

IV. Proposed Methodology

The proposed work in this paper concentrates over the vision based analysis for the recognition of alphabets of the Indian Sign language. The system would be designed to make the computer recognize the alphabets of Indian Sign Language. For the design of the proposed system the first thing required would be the detection of hand from the captured image by the camera.

4.1 Skin Segmentation: Skin Locus

Extraction of Skin Pixels from the color information can be crucial considering the appearance of skin in images is affected by different factors such as illumination, background and camera characteristics. Skin color model should be robust against environmental changes such as changes in the lightning conditions or changes in camera parameters and it should also work with users having different skin colors.

In this paper, we have worked upon analyzing the different skin classification and modeling techniques which are based on color content in the visual spectrum. We have carried out the review and analysis which is divided into three separate parts firstly, we have reviewed the different color spaces that can be used for modeling of skin pixels as shown in Figure 4.1 Secondly, we have studied about the different skin modeling and classification approaches. Nevertheless the above mentioned approaches still have their own limitations owing to the real world conditions that may be illumination and other viewing conditions. Lastly we present our approach that we have used for skin-color detection and adaptation techniques to improve over the skin detection performance in the changing illumination and environmental conditions.

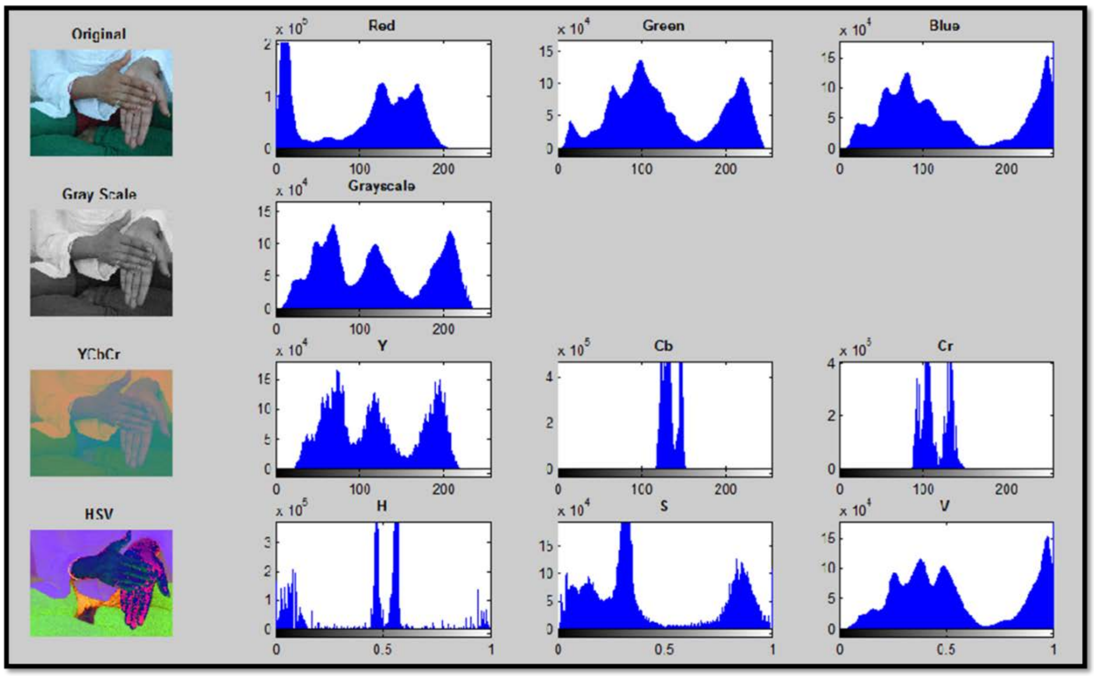


Figure 4.1: Color space modeling for sample skin

Skin detection for the input images for any sign language recognition system plays an important role for the segmentation of the input image. As examined by [11] in the comparative study of different skin chrominance models for various different countries skin tones was done and was found that the most of the skin tones lie in the shell shaped area for the r – g color space that is called as Skin Locus. Soriano[12] proposed the skin locus as illustrated in Figure.4.2 to cope with the changes in the illumination conditions. We have followed the skin locus as proposed by Soriano but with normalized r-g color space with the changes in the thresholds of the skin tones so that the changes in thresholds are suitable for the Indian skin tone. We have employed a skin color distribution obtained empirically to plot the skin locus in the normalized r-g color space. The presence of r in the normalized color space will be deciding over the presence of skin with certain definite thresholds as presence of any threshold of red could not be considered to be presence of skin. Skin locus consists of the lower and upper bounds in form of quadratic equations as given below for g of the color space as the presence of g will be deciding the skin tone, the darker or lighter and the g of upper and lower bound is dependent on specific function of presence of r ,

The lower bound quadratic function for g is

$g = -0.67r2+0.765r+0.1584$

and the upper bound quadratic function is

$g = -1.3852r2 + 1.05r + 0.1652,$

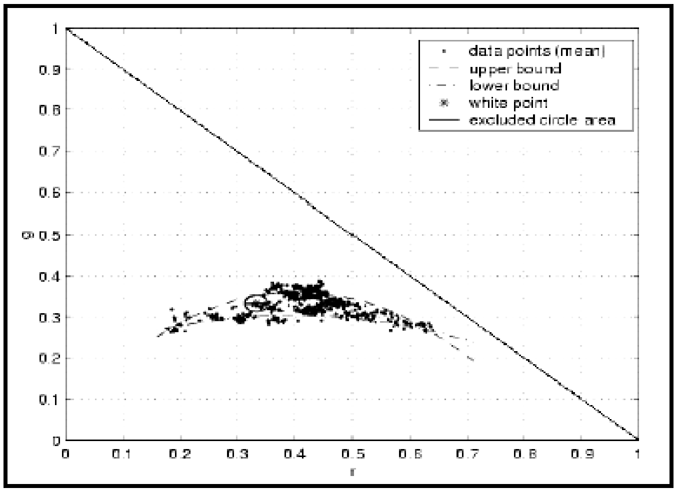


Figure 4. 2 Skin locus of QuickVideo camera[12]

The Figure 4.3 shows the results of skin locus plotted for different Indian skin tones ans has been observed that for darker or lighter skin tones the range of g and r in the r g color space plot varies over a range for g not varying much over the range but it has been found that it is the prence of red in the skin sample at certain threshold that decides that the candidate pixel is skin or not and g only decides about the darker or lighter skin tone.

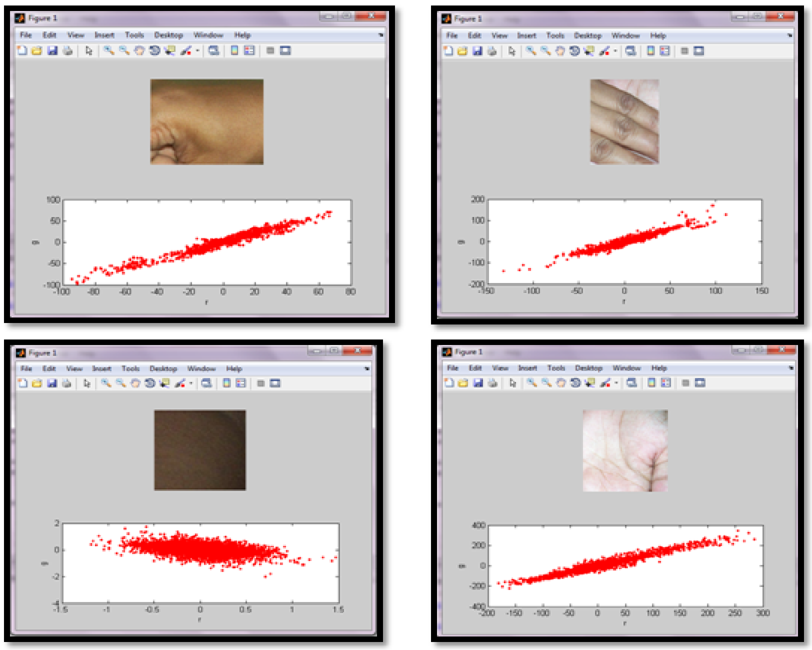


Figure 4.3 Skin Locus obtained for different skin samples

The skin detection results for a few ISL gestures have been depicted and summarized below. The skin color pixels are white and the other ones are black in the skin –segmented image. In our proposed approach of recognizing ISL Gestures we have considered presence of skin as our data to be further processed to achieve our final goal of recognition of signs.

The regions of the skin extracted from the input gestures will act as input to the Feature Extraction scheme of our work. Since our feature extraction method consist of dividing the complete input image to 64 subpatterns which will act as features for the input gesture, thus it becomes essential that detection of skin from the input should be precise for detection of edges also. Importance of edges detection along with the skin segmented image is due to the fact such that the signs which are similar in appearance when considered two dimensionally, the presence of edge will act as feature which can differenciate such similar sign gestures. As can be viewed from Figure 4.4 edges are quite visible even for similar types of signs. Figure 4.4.shows the skin segmented results for a few alphabets of the Indian Sign Language.

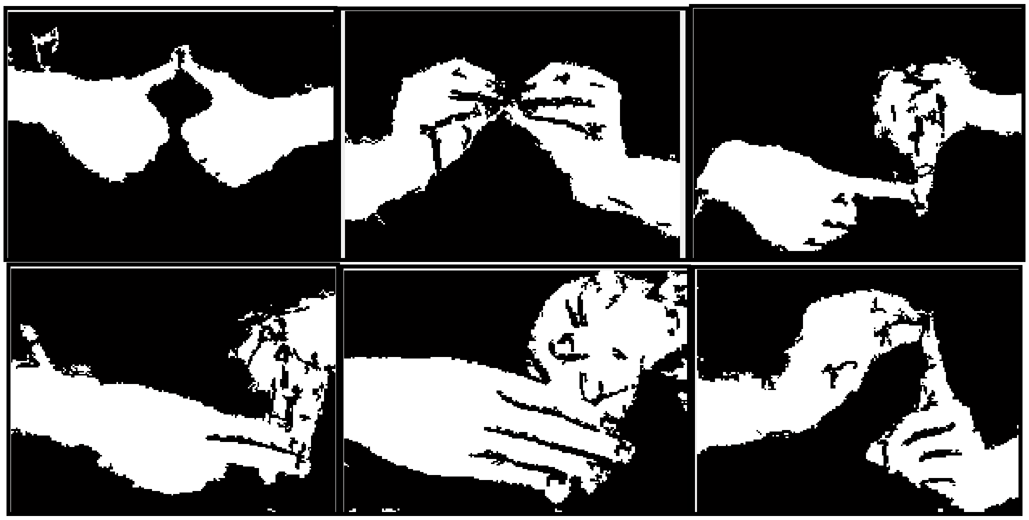
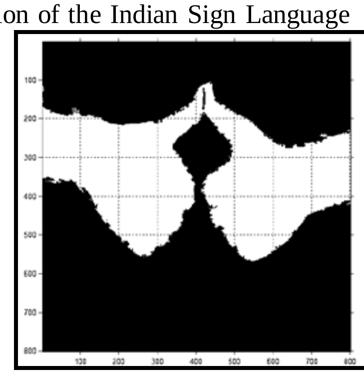


Figure 4.4 Skin detection Results for different ISL signs

The structure of the recognition system is as follows. Initially the image of a sign gesture of ISL is obtained using an i-ball C8.0 web camera. Next stage is the segmentation of the skin pixels after preprocessing and filtering of the captured image. Further after the skin segmented image is obtained it is used to extract the features of the defined gesture for an alphabet of ISL. The extracted features from the segmented image would be used to train as well as test the system for ISL recognition by means of machine learning algorithms.

4.2 Feature extraction

Feature extraction of the skin segmented image is the most important step towards recognition of the Indian Sign Language alphabets. Once the segmented image is obtained the further step is to determine the features from the segmented image. This is important because not the segmented image as a whole might be useful and it might contain some useless or redundant image data. As can be observed from the Figure 4.4 that from the captured image the features corresponding to the particular sign might be contained in the data where the hands are present in the segmented image . Thus feature extraction stage of the recognition system is to represent the segmented image by a set of numerical features that might be according to the particular sign of the Sign language. Also by feature extraction we may reduce the redundant data and thus reducing the dimension of the feature vector.



1	2	3	4	5	6	7	8
9	10	11	12	13	14	15	16
17	18	19	20	21	22	23	24
25	26	27	28	29	30	31	32
33	34	35	36	37	38	39	40
41	42	43	44	45	46	47	48
49	50	51	52	53	54	55	56
57	58	59	60	61	62	63	64

Figure.4.5 A Sample Input divided into its corresponding 64 feature vectors

According to [14] a set of features should fulfill following conditions:

(1) Features of the images belonging to the same class must be similar to each other.

(2) Images belonging to different classes should possess noticeably different features.

(3)Last but most important the Features extracted from the segmented images should be scale, translation, and rotation invariant, which implies that recognition should be obtained irrespective of size, location, and orientation of the sign gesture.

Our method of feature extraction uses the edge information to mark the boundries of the skin and the non skin pixels of the image. Next we find the centroid of the segmented sign followed by the direction of the sign in order to determine the feature vector.

For Feature extraction we make use of the information of border for the sign gesture, the center of the sign gesture and the direction parameter of the sign gesture. The origin is from the center point of the sign from which are considered lengths of the vectors in all directions of the segmented sign.

Consider dx and dy as the coordinates of the center of sign gesture , and px and py be the coordinates for a point on the periphery of the sign gesture then length of the vector will be

$$l_{dp} = \sqrt{(\mathrm{px - dx})^2 + (\mathrm{py - dy})^2}$$

and the direction of the vector as

$$\theta_{dp} = \tan^{-1}\left(\frac{py - dy}{px - dx}\right)$$

Two problems were encountered here. Firstly, how can we determine the useful part of the border, and secondly what should be the appropriate number of vectors to be used? The direction of the sign gesture is utilized in order to find the useful portion from the segmented image.

We have utilized the Adaptive Neuro-Fuzzy Inference Systems (ANFIS)[15] Sugeno type Fuzzy Inference[16] system for the purpose of recognition of Indain Sign Language Alphabet gestures. To construct an ANFIS model for a specific Class includes

• Determination of type of Fuzzy Inference System for the class, and

• Obtaining the Training model using ANFIS by selected Fuzzy Inference System.

Corresponding to each of the 26 alphabets of Indian Sign language an equivalent ANFIS model is built by the training procedure of the ANFIS. Every ANFIS model that is trained produces a output value 1 at the testing phase if the sign gesture presented at the input matches for the defined sign gesture for ISL in the ANFIS trained model. The output 1 corresponds to gesture sign recognized and thus that recognized alphabet is converted to text at the output.

The recognition for the input sign that will be provided by the camera to the trained system will result in 64 outcomes. For the process of recognition from the trained modules for the Signs a voting scheme is followed in order to find that to which class does the input gesture belongs to.

4.3 Effect of Cluster Radius for Training Data on Recognition

It can be observed that the density measure for a data point is a function of the data point's distance to the other data points. Therefore a data point that will be having a number of neighbouring points will possess a higher potential of becoming a cluster center[17]. The cluster radius ra, will be defined as the radius of neighborhood. The points that will be lying outside this radius will exhibit less effect over the density measure of the trained system. Thus the choice of the value of ra, for the training data plays a significant role for the determination of number of clusters. Keeping smaller values of ra, would lead to large number of clusters, whereas larger values of ra, would lead to small number of clusters. The first cluster center is chosen to be the data point that has the highest density measure. Finally the cluster radius affects the complexity of the system in terms of number of rules to be generated for the Fuzzy Inference System as shown in Figure 4.5, 4.6 and 4.7. Larger values of ra would result in less number of clusters and, hence, results in a less number of rules whereas small values of ra means large number of clusters which in turn means larger number of rules generated for the Fuzzy Inference System.

Therefore, value of ra might affect the recognition performance of the system considerably. It is observed that for small values of ra for the training data the recognition rate of the system is very high. But the choice of ra to be very small beyond certain threshold which might depend on other parameters of the system as to no of inputs , no of clusters generated or overlapping of clusters would also lead to misrecognition for any input sign gesture.

The behavior observed for small values of ra is caused by a phenomenon called overfitting[18]. Overfitting is the situation in which the fuzzy system is fitted to the training data so well that its ability to fit to the testing data is no longer satisfactory. In our case, overfitting occurs when the number of rules describing the system is very large, which results in a very specific description of the training data as shown in Figure 4.6.This causes the system to respond very bad for any data that does not fit to that specific description, and therefore, it reduces the system's generalization capability. When the value of ra becomes too large, the small number of generated rules will not be sufficient to convey a good description about the system Figure 4.8. So the behavior will be bad for both training and testing data. Of course, we are not interested in a system with low generalization capabilities. Instead, we are looking for a system that is trained using training data and can perform well with testing data.

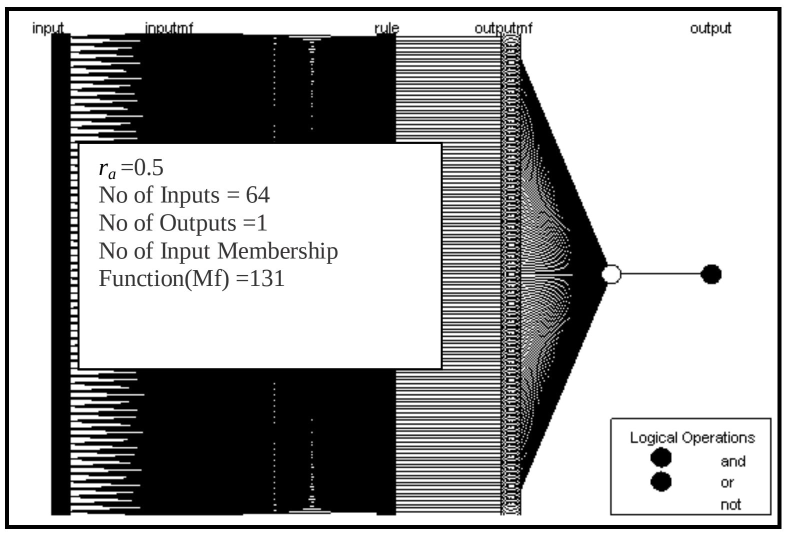


Figure 4.6 Structure of ANFIS with r_a =0.5

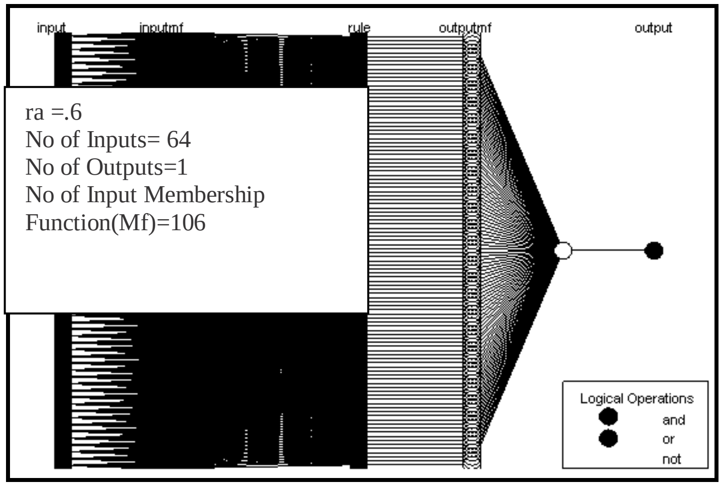


Figure 4.7 Structure of ANFIS with r_a =0.6

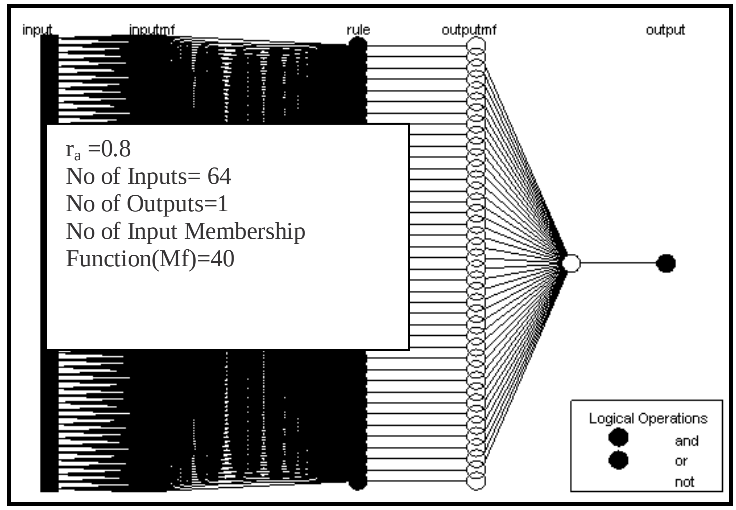


Figure 4.8 Structure of ANFIS with r_a =0.8

V. CREATION OF THE DATABASE FOR OUR WORK

In order to approach the problem of translating signs into text it is necessary to create a database from videos of different signs by multiple signers. Unlike American Sign Language or British sign language, Indian sign language does not have a standard database that is available for use. Hence we have created our own database of Indian signs with reference to Indian Deaf Society [1]. The complete database representation for alphabets of Indian Sign Language is as shown in Figure 1.1 which shows alphabets A to Z for ISL wherein including signs of both single hand and two hand signs.

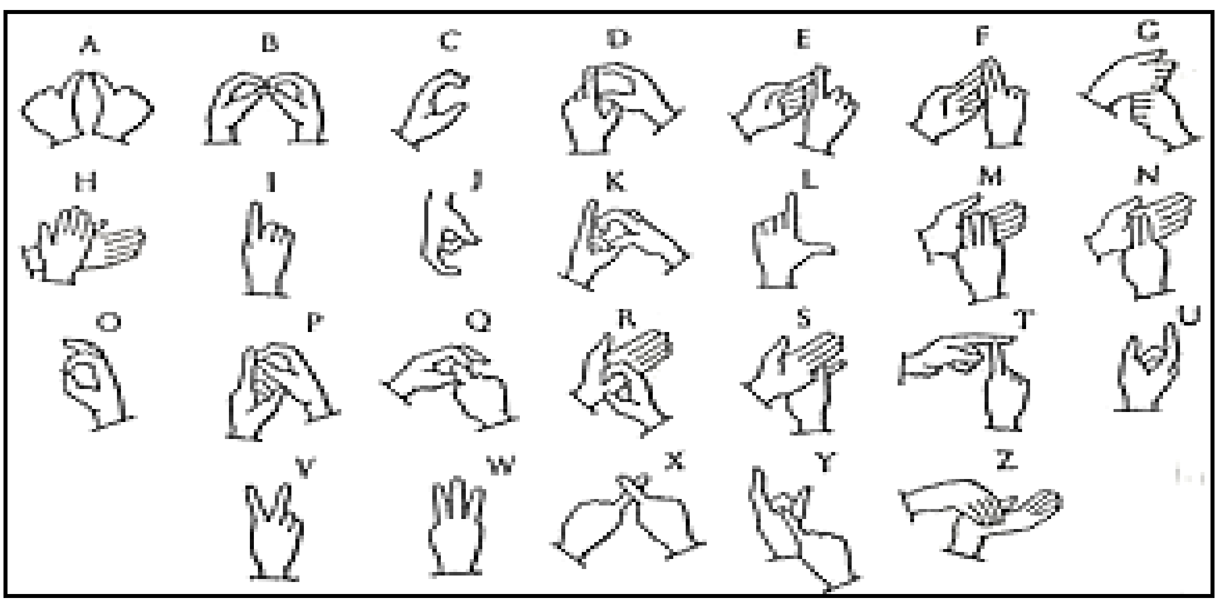


Figure 5.1 Database Representation for ISL[Deaftravel.co.uk]

The data set used for our system for the purpose of training and testing is composed of sign gesture images of ISL for five different datasets of signs of ISL out of the 26 signs shown in Figure 5.1. Samples input corresponding to each sign gesture of ISL were taken from 5 different volunteers. For each signing gesture 100 samples each were used for training purpose.So we had a total of around 2611 training samples for the 26 alphabets of Indian Sign Language for which we obtained different combinations in form of five different datasets in order to test the ability of the system. Training was performed for maximum of 3 epochs reached. The samples were taken from different signers of different age groups for example child or elderly and also for different skin tones. This was basically done in order to have the features extracted for the training datasets to have a variety of inputs. Following are our 5 datasets shown below.

Dataset 1: A, B, C, D, E, G, I, O, U, Q, X (Single and two hand signs)

Dataset 2: F, L, W, V, Y, H, J, N. (Single and two hand signs)

Dataset 3: K, M, T, P, R, S, Z. (Only two hand signs)

Dataset 4:A, X, G, B, D, E, K, T, R, Z. (Ten two hand signs)

Dataset 5: C, I, J, L, O, U, V, W, Y. (Nine Single hand signs)

Here Dataset 1 , Dataset 2 and Dataset 3 combine to recognize all the 26 alphabets of ISL wherein these datasets comprise of both single hand and two hand signs mix in the dataset.

Dataset 4 comprise of ten two hand signs only that is a database having one type of data all of two hand signs . Dataset 5 comprise of 9 single hand signs only that is a database having one type of data all of single hand signs .

The experimental setup consists of a light background and preferably the signer wearing a light colored clothes. This controlled environment reduces tracking and segmentation problems. The RGB videos are acquired using a i-ball C8.0 web camera at a resolution of 1024 X 768 pixels. The images from camera are acquired under normal lighting conditions to simulate real-time environment. For training of system the case of Dataset 1 comprising of 11 signs a total of 4 different signers volunteered where each signer is asked to repeat the sign twenty five times under different conditions such that we provide a total of 100 training samples of each alphabet which turns out to be a total number of 1068 sign sample data for a set of 11 signs of alphabets of Indian Sign Language. The signs in our data base are presented in Figure 5.1 as shown below.

Fig. 5.2 Created Database for our work

VI. Results for Determination of Data Feature Vectors

The feature vector determined for the purpose of recognition of the alphabets is the 1X64 matrix vector obtained by dividing the input gesture image into 64 subblocks as shown in Figure4.5 The feature vector contains values corresponding to the presence of data, for our case data is the presence of skin pixels in the input. The presence of the amount of the skin pixel candidates in each of the 64 subblocks may be different for different signs. There will a numerical value assigned to each vector in the 1X64 matrix as according to amount of the presence of data that in our case is skin data in the particular subblock. The feature vector for two different signs is depicted in form of graphical representation in the Figure 6.1 and Figure 6.2

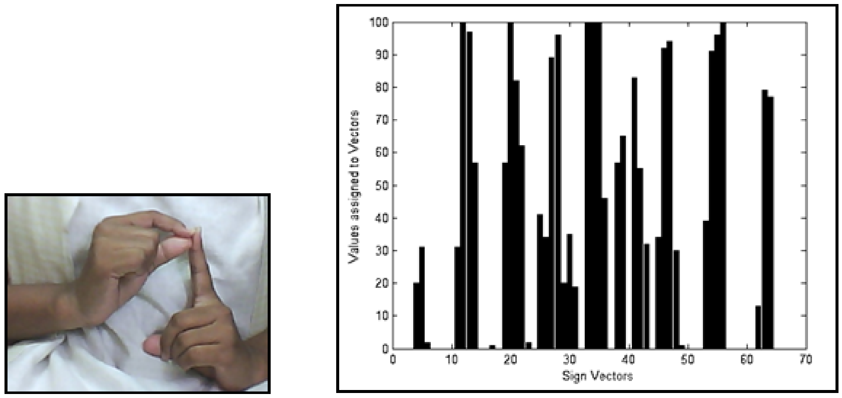


Figure 6.1 Feature Vector Plot for Sign 'P'

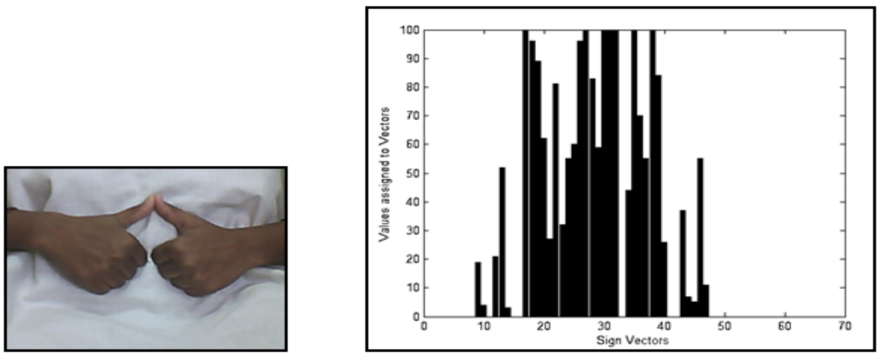


Figure 6.2 Feature Vector Plot for Sign 'A'

6.1Feature Vector of Similar Signs:

From the above graphical representation of two different alphabets it is clear how the feature vectors of different symbols will differ from each other and thus recognition will highly depend on these features.

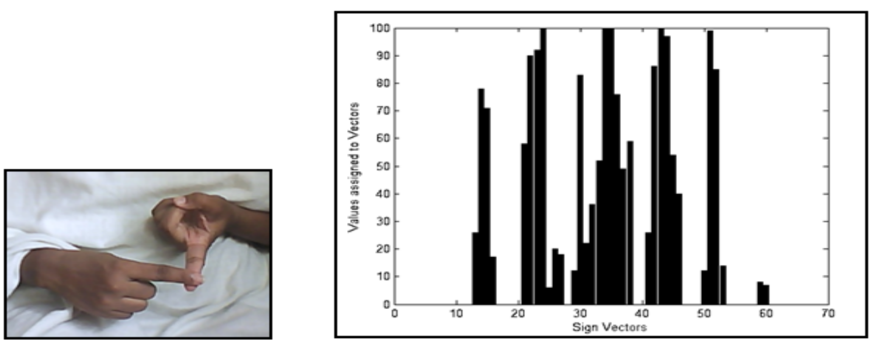


Figure 6.3 Feature Vector Plot for Sign 'E'

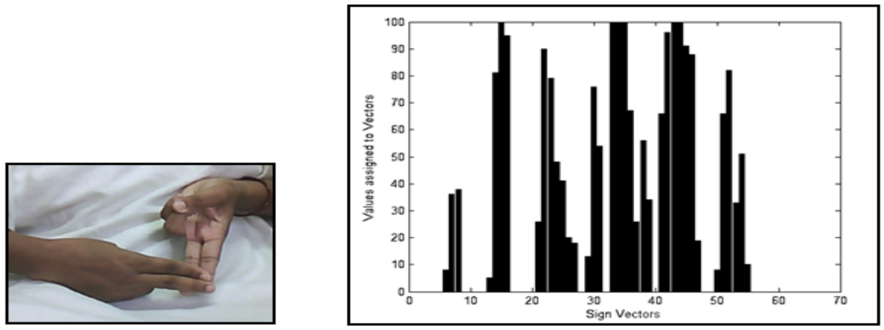


Figure 6.4 Feature Vector Plot for Sign 'F'

From the figures 6.3 and 6.4 it is very clear that how the Feature vector of two similar kinds of signs are quite similar which might affect the recognition of such similar signs.

VII. Recognition of the Signs of ISL Alphabets

Each alphabet of ISL having its unique 64 vectors is used to obtain a training model for the purpose of recognition of alphabets. The actual recognition is obtained by providing the input ISL sign gestures through the webcamera and providing the features of the same after the skin detection and preprocessing stage to the ANFIS trained model. The feature extraction phase will result in 64 different responses. The classification of the input sign gesture by a voting scheme according to the clusters formed out of the training process.

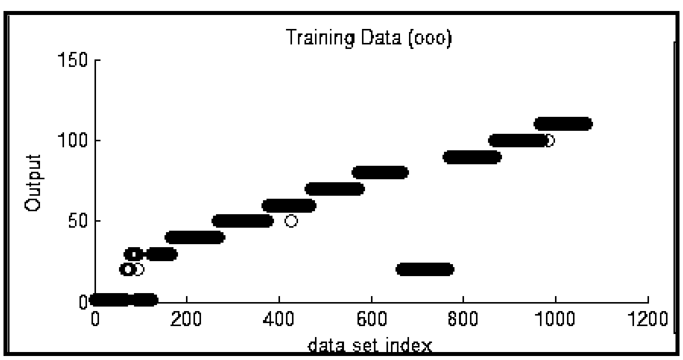


Figure 7.1 Plot of Training Dataset 1

VIII. Results

The performance of our system has been evaluated on the basis of the ability of the system to correctly recognize the alphabets sign gestures to their corresponding input gestures. The recognition rate defined for the system is the ratio of the number of correctly recognized gestures to the total number of input gesture samples. Experiments show that the system can work correctly with sufficient training data, to extend the training database is only a compromise to reduce the lack of training data. Most of the misclassified samples correspond to the gestures that are similar to each other. As an example, Figure 6.3 and Figure 6.3 show the gestures "E" and "F". Because these gestures are similar, their corresponding features are also similar. Therefore, it is probable for a sample of the gesture "E" to be classified as "F" or vice versa. Figure 7.1 shows plots of training for dataset 1 that is for each alphabet approximately 100 samples have been obtained and the same process is repeated for total 11 signs of dataset 1. The datasets 1, 2 and 3 in different combinations comprise of all 26 signs whereas the dataset 4 and 5 are separately for only single handed and only two handed signs.

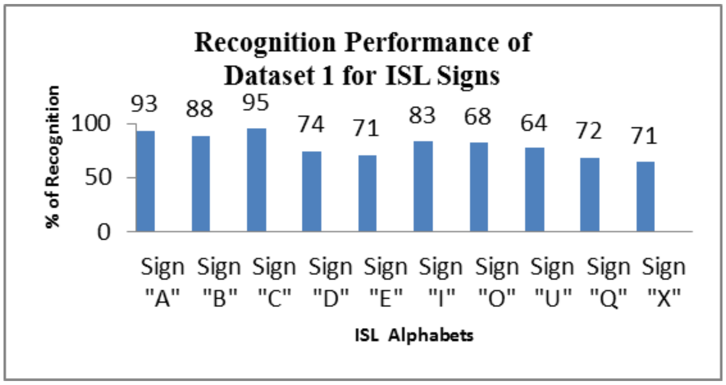
ISL Sign	Test Data	Error	Recognized	% Recognition
"A"	100	7	93	93%
"B"	100	12	88	88%
"C"	100	5	95	95%
"D"	100	26	74	74%
"E"	100	29	71	71%
"I"	100	17	83	83%
"O"	100	18	68	68%
"U"	100	22	64	64%
"Q"	100	32	72	72%
"X"	100	36	71	71%

Table 7.1 The ISL Sign Dataset 1 for Recognition

8.1 Recognition Performance of Dataset 1 for ISL Signs

For dataset 1 that is the 11 alphabets "A, B, C, D, E, G, I, O, U, Q, X" which comprise of both Single and two hand signs , thus the recognition results in this case would exhibit the capability of the recognition system to be able classify for both types of signs in a single training dataset. It has been observed that most of the misclassification results for were for the alphabets "O" and "U" due to similarity in the features of these two alphabets.

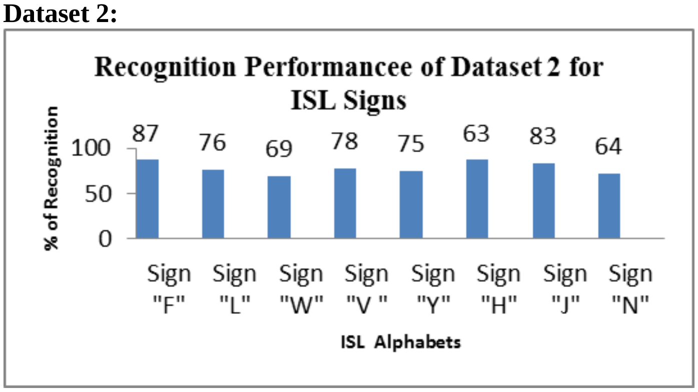
Dataset 1:



8.2 Recognition Performance of Dataset 2 for ISL Signs

For dataset 2 that is the 8 alphabets “F, L, W, V, Y, H, J, N” which comprise of both Single and two hand signs , thus the recognition results again in this case would exhibit the capability of the recognition system to be able classify for both types of signs in a single training dataset. For dataset 2 it has been observed that most of the misclassification results for were for the alphabets “H” and “N” due to similarity in the features of these two alphabets.

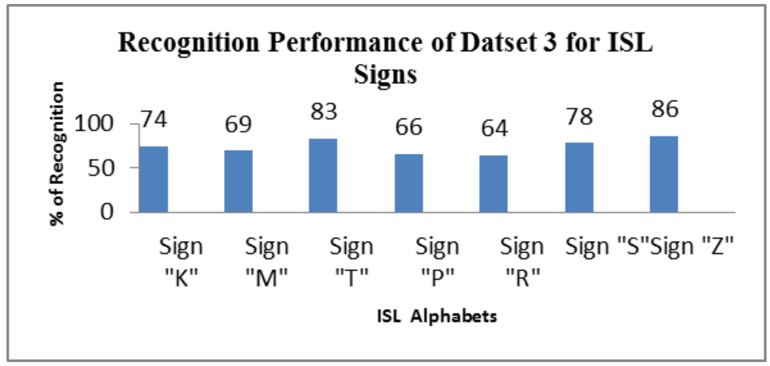
Dataset 2:



8.3 Recognition Performance of Dataset 3 for ISL Signs

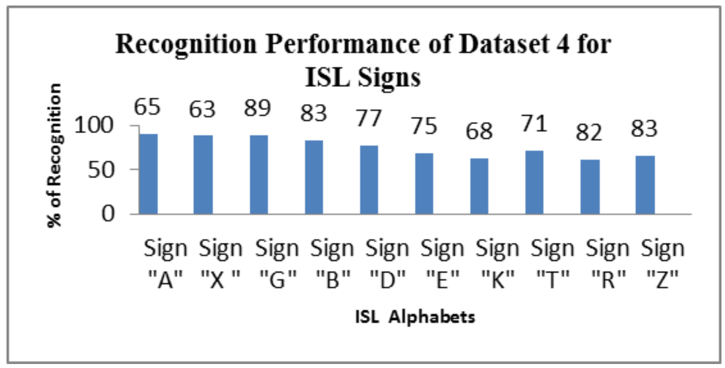
For dataset 3 that is the 7 alphabets “K, M, T, P, R, S, Z” which comprise of only two hand signs, thus the recognition results in this case would exhibit the capability of the recognition system to be able classify for all two handed signs in a single training dataset. For dataset 3 it has been observed that most of the misclassification results for were for the alphabets “K” and “P” due to similarity in the features of these two alphabets.

Dataset 3:



8.4 Recognition Performance of Dataset 4 for ISL Signs

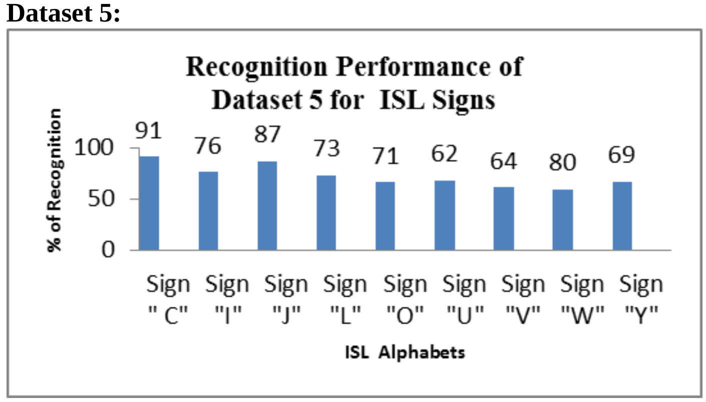
Dataset 4:



For dataset 4 that is the 10 alphabets “A, X, G, B, D, E, K, T, R, Z” which comprise of only two hand signs, thus the recognition results in this case would exhibit the capability of the recognition system to be able classify for all two handed signs in a single training dataset.For dataset 1 it has been observed that most of the misclassification results for were for the alphabets “A” and “X” due to similarity in the features of these two alphabets.

8.5 Recognition Performance of Dataset 5 for ISL Signs

Dataset 5:



For dataset 5 that is the 9 alphabets “C, I, J, L, O, U, V, W, Y” which comprise of only single hand signs, thus the recognition results in this case would exhibit the capability of the recognition system to be able classify for only single handed signs in a single training dataset

For dataset 5 it has been observed that most of the misclassification results for were for the alphabets “O” and “U” due to similarity in the features of these two alphabets.

IX. Conclusion

Having the limitation of providing the training data for all 26 alphabets as a whole because of the more no of feature vectors and the bulky size of the training sample data , we have provided the training for the data samples in five different datasets. Each dataset is different in its own way in terms of type of alphabets included in each dataset. For example the five datasets include in variety of combinations of the ISL alphabets with only single hand signs in some or only two hands signs in some or a mixture of both type of signs in the other. But whatever combinations we might have considered we have been able to recognize for all 26 alphabets of Indian Sign Language. Most of the misclassified samples correspond to the gestures that are similar to each other. As an example, shows the gestures “E” and “F”. Because these gestures are similar, their corresponding features are also similar. Therefore, it is probable for a sample of the gesture “E” to be classified as “F” or vice versa.

References

[1] Indian Sign language, empowering the deaf, Indian Sign Language Database
<http://www.deafsigns.org>.

[2] Ulrich von Agris, Daniel Schneider, Jorg Zieren, and Karl-Friedrich Kraiss. “Rapid signer adaptation for isolated sign language recognition.” In CVPRW ’06: Proceedings of the2006 Conference on Computer Vision and Pattern Recognition Workshop, page 159, IEEE ComputerSociety Washington, DC, USA, 2006.

[3] C.-F. Juang and Ksuan-Chun Ku. “A recurrent fuzzy network for fuzzy temporal sequence processing and gesture recognition.Systems” Man, and Cybernetics, Part B: Cybernetics, IEEETransactions on, 35(4):646–658, Aug. 2005.

[4] Gaolin Fang, Wen Gao, and Debin Zhao. “Large vocabulary sign language recognition based on hierarchical decision trees.”Proceedings of the 5th international conference on Multimodal interfaces - ICMI ’03, page 125, 2003.

[5] Byung-woo min, Ho-sub yoon, Jung soh, Takeshi ohashi and Toshiaki jima”, Visual Recognition of Static/Dynamic Gesture: Gesture-Driven Editing System”, Journal of Visual Languages & Computing Volume10,Issue3, June 1999, Pages 291-309.

[6] Kouichi Murakami and Hitomi Taguchi, Gesture Recognition using Recurrent Neural Networks, in CHI ‘91 Proceedings, pp. 237-242 (1991).

[7] T. Shanableh, K. Assaleh,Arabic sign language recognition in userindependent mode”, IEEE International Conference on Intelligent and Advanced Systems 2007.

[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),Vol. 4 No. 03 March 2012 pp 406-415

[9] P.Subha Rajam, G. Balakrishnan, "Indian sign language recognition system to aid deaf-dumb people", IEEE International Conference on Computing Communication and Networking Technologies (ICCCNT), Page(s): 1-9, Trichy, India, 2010.

[10] R. Pravin Futane, Rajiv V. Dharaskar, "Hasta Mudra": An interpretation of Indian sign hand gestures" 2011 IEEE 3rd International Conference on Electronics Computer Technology (ICECT), Vol:2, Page(s): 377-380, Kanyakumari , India ,2011

[11] J.C. Terrilon, M.N. Shirazi, H. Fukumachi, and S. Akamatsu, “Comparative performance of different skin chrominance models and chrominance spaces for the automatic detection of human face in color images,” Proc. 4th IEEE International conforence on Auto-matic Face and Gesture Recognition, pp.54-61, Grenoble,France, March, 2000.

[12] M. Soriano, B. Martinkauppi, S. Huovinen, and M. Laaksonen, “Using the skin locus to cope with changing illumination condition in color-based face tracking,” Proc. IEEE Nordic Signal Processing Symposium, pp.383-386, Kolmarden, Sweden, 2000.

[13] ALBIOL, A. TORRES, L., AND DELP, E. J. 2001. Optimum color spaces for skin detection. In Proceedings of the International Conference on Image Processing, vol. 1, 122.

[14] Jos L. Hernndez-Rebollar, Robert V. Lindeman, and Nicholas Kyriakopoulos. “A multi-class pattern recognition system for practical finger spelling translation” In Proceedings of the 4th IEEE International Conference on Multimodal Interfaces,pages 185-190, 2002.

[15] J.-S.R. Jang, ANFIS: Adaptive-Network-Based Fuzzy Inference System, IEEE Trans. Systems Man Cybernet. 23 (1993) 665–685

[16] Takagi T, Sugeno M, 1985. Fuzzy identification of systems and its applications to modeling and control. IEEE Transactions on Systems, Man, and Cybernetics, 15: 116–132.

[17] Sin, S. K., and De Figueiredo. 1993. Fuzzy System Designing Through Fuzzy Clustering and Optimal preDefuzzification. Proc. IEEE International Conference on Fuzzy Systems. 190-195

[18] .S.R. Jang, C.-T. Sun, E. Mizutani, Neuro-Fuzzy and Soft Computing, Prentice Hall, Englewood Cliffs, NJ, 1997

Authors

First Author – Shweta Dour **-** She received her BE in Electronics and Communication Engineering from Rajasthan University , Jaipur. She received her ME in Electronics and Telecommunication Engineering from Mumbai University in 2012. Presently a research scholar at Bhagwant University , Ajmer and working over Real time recognition of Indian Sign Language in the field of Computer vision and Pattern Recognition.

Second Author – M.M. Sharma- He received his Phd degree from Malaviya National Institute of Technology, Jaipur. He pursued his B.E (Electrical Engineering) from Regional Engineering College, Srinagar, Kashmir (J&K)and M.Tech (Computer Technology) from IIT Delhi. He is Presently Honorary Secretary IEEE MTTS India Council Chapter & Principal Govt Engineering College Jhalawar & Principal, Govt. Engineering College,Ajmer