Indian Sign Language Character Recognition using Neural Networks

Padmavathi . S
Asst.Professor
Dept. of Information Technology
Amrita Vishwa Vidyapeetham
Coimbatore, India

Saipreethy.M.S
Dept. of Information Technology
Amrita Vishwa Vidyapeetham
Coimbatore. India

Valliammai.V
Dept. of Informayion Technology
Amrita Vishwa Vidyapeetham
Coimbatore, India

ABSTRACT
Deaf and dumb people communicate among themselves using sign languages, but they find it difficult to expose themselves to the outside world. This paper proposes a method to convert the Indian Sign Language (ISL) hand gestures into appropriate text message. In this paper the hand gestures corresponding to ISL English alphabets are captured through a webcam. In the captured frames the hand is segmented and the neural network is used to recognize the alphabet. The features such as angle made between fingers, number of fingers that are fully opened, fully closed or semi closed and identification of each finger are used as input to the neural network. Experimentation done for single hand alphabets and the results are summarized.

General Terms
Neural Network Algorithms, Hand Gesture Recognition.

Keywords
Indian Sign Language Recognition, Hand gesture recognition, neural networks, activation function.

1. INTRODUCTION
Communication is defined as exchange of thoughts and messages either by speech or visuals, signals or behavior. Deaf and dumb people use their hands to express their ideas. The gestures include the formation of English alphabets. This is called sign language. When they communicate through the computer; the gestures may not be comfortable for the person on the other side. Therefore for them to understand easily, these gestures can be converted to messages. The sign language is regional. There is a separate sign language in America that uses only one hand for picturing the gestures. But the Indian sign language is totally different. It uses both the hands for representing the alphabets. While there are lot of efforts going into American Sign Language detection the same cannot be said about Indian Sign Language. The existing systems concentrate on general gesture recognition of hand gestures for any action. This paper proposes recognition of Standard Indian sign language gestures. Unlike the conventional method for hand gesture recognition technique which makes use of gloves or markers or any other devices, the method proposed in this paper does not require any additional hardware and makes the user comfortable. This paper uses the hand gestures captured through webcam as the input. Image processing techniques are being used to extract the features which are used to identify the alphabets using neural networks.

The input image is segmented for processing. After segmentation, the regions like fingers and palm are extracted. Using this extracted region the fingers are identified. The different features of the fingers like angle, posture of the finger are detected. The posture of the finger specifies whether the finger is semi closed, half closed, fully closed or fully open. Feature like whether the hand shown is right or left is also found. These features are stored as a vector. These vectors are then used as input for recognizing the correct gesture using neural network. Different architectures of neural networks are analyzed. This paper is organized into the following sections. Section 2 talks about the various techniques available for gesture recognition. Section 3 deals with proposed method used for recognition of gestures using neural networks. Section 4 contains the results obtained on experimenting various inputs on 3 different neural network architectures. Conclusion and future work is given in the section 5.

2. LITERATURE SURVEY
Extraction of hand region plays a major role in hand gesture recognition. Skin color based segmentation technique is widely used for segmenting hands, faces etc. These techniques rely upon the color model used for segmentation[6]. Various extraction procedures such as using threshold values, combining collection of low level information to high level feature information, projecting the object using Eigen vectors, detecting finger tips by segmentation, using the concept of kinematics and dynamics of the body are discussed in [2]. Some feature extraction techniques like finger tip detection, finger length detection are discussed in[6]. Many learning algorithms can be used for recognition of gestures. SVM based method is discussed in [5].A Hidden Markov model based recognition is discussed in [3]. A neural network based approach is discussed in [4]. Feed back or Back Propagation of the output is what enables learning in a neural network. One of the first systems to use neural networks in hand posture and gesture recognition was developed by Murakami. Hand postures were recognized with back propagation neural network that had 3 layers, 13 input nodes, 100 hidden nodes, and 42 output nodes, one for each posture to be recognized. The method in [11] achieved 77% accuracy with an initial training set. An increase in the accuracy to 98% in [12] is reported for participants in the original training set when the number of training patterns was increased from 42 to 206.

Symeonidis [13] trained a single perceptron back propagation model to recognize hand gestures of American Sign Language. An accuracy of 78% was reported on an overall basis[13].

3. PROPOSED METHODOLOGY

The alphabets in Indian Sign Language shown in Figure1, uses both the hands which differentiate it from American Sign Language. Indian Sign Language characters are almost similar to the characters themselves. The alphabets could be characterized by the angle between the fingers, which finger is open, how much the particular finger is open and how many fingers are open.

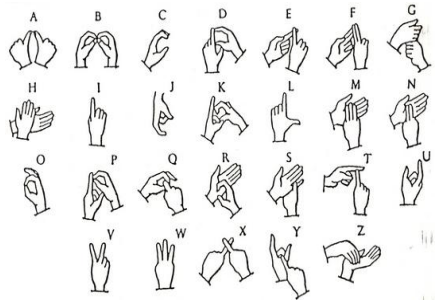

Fig 1: Indian Sign language – Characters[8]

The block diagram shown in Fig 2 explains clearly about the phases involved in the process. The webcam is used to capture the gesture made by the person in front of the computer. The input video is converted into frames and HSI color model based segmentation is applied on each frame. After segmentation, in feature extraction, certain features of the hand like centroid of the hand helps identify features like posture of the fingers and angle between them. After the phase of extraction, is the recognition phase where the extracted features are fed into the neural network to recognize the particular character.

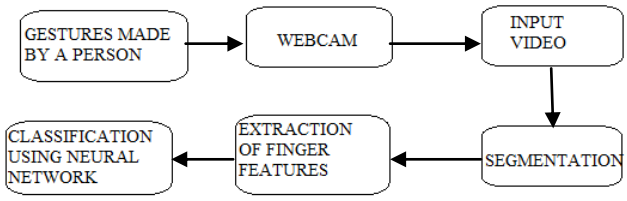

Fig 2: Overall Flow of The System

3.1 Segmentation

Since the segmentation is done using skin color model, effect of luminosity should be segregated from the color components. This makes HSI color model a better choice than RGB. The optimal Hue and saturation values for hand as specified in [6] is H<25 or H>230 and S<25 or S>230.

The input frames in RGB model are converted to HSI model using Eq(1). One assumption made is that the hand is the largest skin color object in the input image. After segmenting, the hand region is assigned a white color and other areas are assigned black.

$$h = 2r - \{0.5[(r-g)+(r-b)] / (r-g)^2 + (r-b)(g-b)^{1/2}\}$$

$$s = 1 - 3\min(r,g,b) \qquad \text{Eq(1)}$$

where r=R/R+G+B, b=B/R+G+B and g=G/R+G+B, h is the hue component and s is the saturation component, R, G and B represents the red green and blue component of the pixel respectively.

3.2 Feature Extraction

Distance transform method [6] is used to identify the centroid of the hand.

Depending on the centroid and the segmented hand region an appropriate size for the structuring element (a disc) is calculated. The segmented image is eroded with the structure element as given in Eq(2). This results in palm region.

$$A \ominus B = \bigcap_{z} d\, E \mid B_z \subseteq A \qquad \text{Eq(2)}$$

where A is the binary image in E, B is the structuring element and Bz is the translation vector of B at pixel z.

To avoid sharp edges in the extracted palm region dilation as specified in Eq(3) is applied.

$$A \oplus B = \bigcap_{z} z \mid E \mid \hat{B}_z \cap A \neq Q \qquad \text{Eq(3)}$$

where A is the binary image in E, B is the structuring element and (Bsz) is the symmetric of B at pixel z.

The palm region so obtained is subtracted from the segmented hand. This leaves the fingers and the wrist portion as disconnected regions. If each region has an area within a predefined range, then it is identified as finger region. This eliminates the wrist region from being considered for gesture recognition. The number of such regions gives the number of fingers that are open. For each of the finger region extracted the farthest point from and closest point to the palm region along the major axis is identified as the finger tip and the base point respectively.

With the base points of each finger known, the distance between every possible pair is calculated. Since the distance between the base point of thumb with any other finger is always larger, this criteria is used to identify the thumb. With all the fingers opened and widely spread the distance of each open finger to every other finger are found. These values are used as thresholds for identifying the fingers that are opened similarly the length of each finger calculated initially are used for identifying whether a finger is semi closed.

To account for the semi closed fingers, each finger except the thumb is divided into three parts using Eq(4) and Eq(5).

$$p = 1/3 * \text{major axis1ength} \qquad \text{Eq(4)}$$

$$q = 2/3 * \text{major axis length} \qquad \text{Eq(5)}$$

The angle made by the major axis of each finger with respect to the reference line passing through the centroid is measured with all fingers open and widely spread. During the experimental phase the angle made by each finger with the horizontal line passing through the centroid is calculated and then the angle between the fingers is calculated as difference of the corresponding angles.

3.3 Artificial Neural Network

An ANN is used to recognize single hand alphabets viz L, C, I, U, V and W. The ANN learns to recognize these alphabets by adjusting the weights associated with the neurons. The features extracted from the previous phase are used as input for the artificial neural network. The neural network used in this paper is a back propagation neural network which has 1 input layer, 2 hidden layers and 1 output layer as shown in Fig 3.

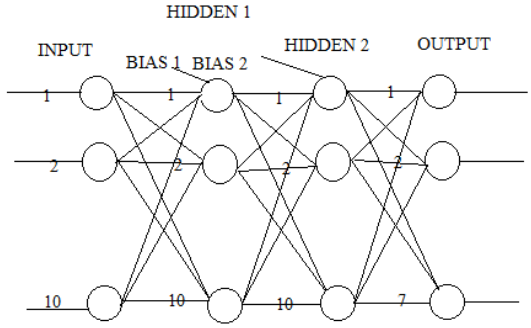

Fig 3: Proposed Back Propagation Neural Network Architecture

The output of a neuron depends on the activation function, the input and the weights of the connections. If X represents the input vector, W represents the weight vector and F represents the activation function, then the output Yj of a neuron 'j' is computed as in Eq. 6 and Eq. 7. In the presence of a bias the output is calculated as shown in Eq. 8. The output of one layer is passed as input to the next layer and the calculation in a similar way continues up to the output layer.

$$I = X.W \quad \text{Eq(6)}$$

$$Y = F(I) \quad \text{Eq(7)}$$

$$Y = F(I + b) \quad \text{Eq(8)}$$

The output vector so obtained from the output layer is used to recognize the character in the testing phase. In the training phase the weights are assigned to some random values initially which has to be adjusted according to the error in the calculated output. The weights are updated as in Eq. 9. The desired output which is the sign language character is known in the training phase and hence the error between the calculated output and the desired output is used to adjust the weights of the neurons.

$$W = W' + REX \quad \text{Eq(9)}$$

Where R is the Levenberg-Marquardt Learning Rule and Ej is the error term which is calculated using Eq. 10.

$$E = Y(1 - Y).(D - Y) \quad \text{Eq(10)}$$

Where Dj is the desired output.

In the back propagation network the error terms are propagated backwards until input layer and the weights are adjusted at each level. For the previous layers, the weights are adjusted using Eq. 9 with error term modified as shown in Eq. 11.

$$e = Y.(1 - Y)./ e.W \quad \text{Eq(11)}$$

Where ek is the error term of each neuron 'k' in the layer that is successive to the layer being considered. The weight adjustment also called as training of the neural network is done repeatedly until the stopping criterion is reached. The stopping criterion depends on the error term or the repeated number of input sets generally called as epochs.

Various activation function such as sigmoid, tanh and hard limiting functions as given in Eq 12, Eq 13 and Eq 14 respectively could be used for the neural network. Their corresponding graphical representations are shown in Fig 4, Fig 5 and Fig 6 respectively.

$$sig(x) = \frac{1}{(1 + e^{-x})} \quad \text{Eq(12)}$$

$$\tan(x) = \frac{2}{(1 + e^{-2x})} - 1 \quad \text{Eq(13)}$$

$$hardlim(x) = 1 if x \$ 0 \quad \text{Eq(14)}$$

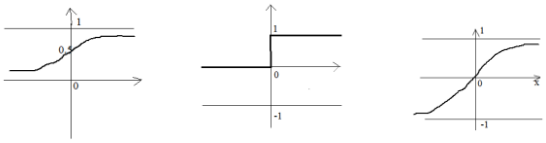

Fig 4: Sigmoid **Fig 5: Hardlimiting** **Fig 6: Tanh**

4. EXPERIMENTAL RESULTS

The ISL characters use both the hands, some fully open, some semi closed and some half closed. The difference in these features between the alphabets was identified and was categorized into groups accordingly. One major category depends on the number of hands used. The alphabets using one hand in the Indian Sign Language and the features required for their recognition are shown in Table 1. The angles represented are an approximate estimation as the actual values vary at runtime due to the change in hand position of the person showing the gestures.

Table 1 – Features of Alphabets Using Single Hand

Alphabet	Fingers opened	Number of Fingers	Angle Between	Posture
C	Thumb and Index	Two	45	Index-Semi Closed
I	Index	One	0	Fully opened
L	Thumb and Index	Two	40	Fully opened
U	Thumb and	Two	60	Index-semi

	Index			closed
V	Middle and Index	Two	20	Fully closed
W	Index, middle and ring	Three	14(between index and middle),20(between middle and ring)	Fully opened

The first 50 frames of each input video are utilized to extract the finger details in which all the five fingers are kept opened. The length and the angle between them are found and stored.. The remaining frames are used for generating the feature vector relative to these details. This makes the features to be independent of the nature of hand and the user. Fig 7 represents an input video frame used during the extraction phase.

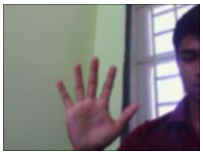

Fig 7: Input Image

Fig 8: Binary Image

The segmentation is performed on this input image and a binary image is obtained as in Fig8. Distance transform is applied on the binary image and the centroid is found as shown in Fig 9.

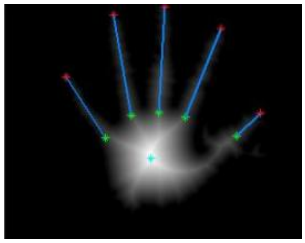

Fig 9: Image after distance transform

After obtaining the finger region, the finger tips and length of each finger is found as explained in section 3. Then each finger's length is divided into 3 parts as shown in fig 10. In the Figures the red points indicate the finger tips, the lines indicate each finger region and the lowest green point represents the base point and the other two green points represent p and q sections as explained in section 3. Using the base points of each finger and the centroid, the angles are calculated as explained in section 3. These extracted angles and length of the fingers are used as the input feature vector for the neural network.

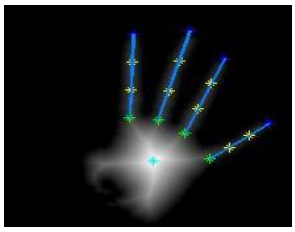

Fig 10: Four Fingers with the partition of three in each

The back propagation neural network used in this paper has 1 input layer, 2 hidden layers with 10 neurons each and 1 output layer with 7 neurons. A 10-tuple feature vector is used for training and testing. First 5 represent the posture of the fingers, next 4 represents the angle between the five fingers and the last is to indicate the hand used in making the gesture. The value of each posture is indicated as 0, 1, or 2 depending on whether the finger is fully closed or semi-closed or fully open respectively. A value 1 or 0 representing hand in the feature vector indicates the right hand and the left hand respectively. The output vector consists of seven classes which represent L, C, I, U, W, V and nothing which represents class whose features does not belong to any of the characters. This includes gestures with fully opened four fingers, fully opened five fingers, thumb finger opened and so on.

Experimentation is done for different number of hidden layers, different number of neurons in hidden layers and various transfer functions while keeping the input and output layer fixed as specified earlier. The stopping criterion is taken as 100 epochs. Tanh and sigmoid functions performed well. The input feature vector was fed to different networks like feed forward, cascade feed forward and the proposed neural network and their results were analyzed. To compare the performance of the network, each alphabet is recorded separately in 7 videos consisting of 150 frames each with gestures made by 2 persons. 80 videos in total are used for training. Gestures made by 3 other persons for each alphabet is recorded and used for testing. The feature vector is extracted from the last 100 frames in each video. The confusion matrix for feed forward neural network with one hidden layer consisting of 10 neurons is shown in Table 2.

Table 2: Confusion matrix of feed forward network

	L	C	I	U	V	W	Nothing
L	100						
C	77						23
I		2	65	1	3		29
U		4		92			4
V					2	85	13
W					4	96	
Nothing		3	82	2			13

The character C was misclassified as L and V was misclassified as W. Classification of I was not proper as evident from the table. The confusion matrix for cascade feed

forward neural network with one hidden layer consisting of 10 neurons is shown in Table 3. Though the recognition of C and V has improved greatly, most of I and V were not identified as characters. The confusion matrix for proposed back propagation neural network with 2 hidden layers is given in Table 4.

Table 3: Confusion matrix of cascade feed forward network

	L	C	I	U	V	W	Nothing
L	100						
C		100					
I		1	62	6		2	29
U		5		87			8
V					77		23
W						100	
Nothing		3	4				93

Table 4: Confusion matrix of the proposed neural network

	L	C	I	U	V	W	Nothing
L	100						
C		65					35
I	4	3	72	2	3		16
U		4		92			4
V					87	13	
W						100	
Nothing		2	3				95

Though the recognition of C has decreased, the recognition of I, U, V have increased to a greater extent. When the number of neurons was increased to 15 and 20, the performance of the neural networks remained the same except feed forward network. The results of feed forward network obtained when 25neurons were used in the hidden layer is given in Table 5.

Table 5: Confusion matrix of feed forward network with 25 hidden neurons

	L	C	I	U	V	W	Nothing
L	100						
C		77					23
I	4	3	61	31			1
U		10		87			3
V				1	96		3
W					4	94	2
Nothing			4			7	89

In all the confusion matrices given above, the classes are sometimes classified as nothing which is due to low illumination. With improper illumination some other skin colored objects in the background is also taken into account and hence classified as nothing. In general for the gestures made through web cam the proposed method had the precision, recall and specificity as 89.47%, 89.78% and 97.54% respectively. When experimented under good illumination, the segmentation and hence the performance improved to greater extent. A comparison between different networks under good illumination is summarized in Table 6.

Table 6: Comparison of accuracy with good illumination

LETTER	D [14]	FF	CFF	BPN
L	80	97	95	100
C	42	93	95	97
I	88	100	100	100
U	60	87	82	96
W	71	100	100	100
V	85	100	100	100
Nothing	72	100	100	100

The first column in Table 6 shows the single hand alphabet, the second specifies the percentage of accuracy obtained by considering the distance between the feature vectors, third fourth and fifth columns specify the percentage of accuracies obtained by Feed Forward(FF), Cascade Feed Forward(CFF) and the proposed Back Propagation Network(BPN) respectively. From the table it is clear that the proposed back propagation neural network has the highest accuracy with 99%, precision 89.47%, recall 89.78% and specificity 97.54%.

5. CONCLUSION AND FUTURE WORK

The Indian sign language alphabets could be identified from the input hand gesture video by identifying the fingers and their postures. The segmentation of the hand and the fingers play a crucial role in such process. The features of the English characters have been obtained for single hand alphabets and were classified using neural networks. Accuracy was increased when neural networks were used. In our case proposed neural networks with sigmoid transfer function gave better results compared to other architectures. Due to segmentation problem the features of letters C and U were mixed up. The effect of illumination also adds up to the improper segmentation. The implementation of segmentation

for both the hands is in progress and its robustness varies when there is an overlap of hands. The detection capability of the system could be expanded to body gestures as well.

6. ACKNOWLEDGMENTS

We would like to convey our thanks to our project mates Ajith.J, Niranjan.M, Karthik.K.S and Shangeetha.R.K for helping us in reaching our goal. We would like to extend our thanks to our seniors for helping us in our project. We also extend our thanks to all the faculty members and non-teaching staff of our college, our parents and our friends whose timely help and co-operation is worth mentioning.

7. REFERENCES

[1] Justin K. Chen, Debabrata Sengupta, Rukmani Ravi Sundaram Vaishali.S.Kulkarni et al., "Sign Language Gesture Recognition with Unsupervised Feature Learning", (IJCSE) International Journal on Computer Science and Engineering[Vol.02,No. 03,560-565] ,2010.

[2] Vaishali S. Kulkarni, Dr. S.D.Lokhande, "Appearance Based Recognition of American Sign Language Using Gesture Segmentation", [Online].

[3] Mahmoud Elmezain, Ayoub Al-Hamadi, Bernd Michaelis "A Real-Time Capable System for Hand Gesture Recognition Using Hidden Markov Models in Stereo Colors Image SEq.ences", (IESK) Otto-von-Guericke-University Magdeburg Germany.

[4] Klimis Symeonidis, "Hand Gesture Recognition Using Neural Networks", August 23, 2000.

[5] Ilan Steinberg, Tomer M. London, Dotan Di Castro, "Hand Gesture Recognition in Images and Video", [Online].

[6] R.M. Arunachalam, M. Ashok Gowtham, and R. Aarthi ,"Identifying Fingertips for Human Computer Interaction", ISSN: 2079-2557 Science Academy Publisher, United Kingdom www.sciacademypublisher.com 1375, International Journal of Research and Reviews in Computer Science (IJRRCS) ,[Vol. 3, No. 1],.February 2012.

[7] Ravikiran J, Kavi Mahesh, Suhas Mahishi, Dheeraj R,Sudheender S, Nitin V Pujari, "Finger Detection for Sign Language Recognition", International Multi Conference of Engineers and Computer Scientists, Hong Kong, IIMECS 2009,March 2009.

[8] http://www.deaftravel.co.uk/signprint.php?id=27

[9] Henrik Jonsson, "Vision-based segmentation of hand regions for purpose of tracking gestures", MSc thesis, Department of Computer Science, Umea University, Sweden, December 2008.

[10] Rosalyn R Porle, Ali Chekima, Farrah Wong, G Sainarayanan, "Performance of Histogram-Based Skin Colour Segmentation for Arms Detection in Human Motion Analysis Application", published in World Academy Of Science, Engineering And Technology, Issue 28 April 2009, v52 page no 1269-1274.

[11] Murakami, Kouichi, and Hitomi Taguchi. "Gesture Recognition Using Recurrent Neural Networks." In Proceedings of CHI'91 Human Factors in Computing Systems,237-242, 1991.

[12] Russell, Stuart, and Peter Norvig. Artificial Intelligence: A Modern Approach. Prentice Hall,NJ, 1995.

[13] Klimis Symeonidis," Hand Gesture Recognition Using Neural Networks", August 2000.

[14] R.K.Shangeetha,V.Valliammai,.Padmavathy, "Computer vision based approach for Indian sign language character recognition" in MVIP – December 2012.