Indian Sign Language Recognition System

Kumara Maruthi M. Mullur

Promantia Global Consulting L.L.P,
Bangalore, Karnataka, India

***Abstract* - Indian Sign Language (ISL) is a language which, instead of acoustically conveyed sound patterns, uses gesture language to convey meaning, which was developed for those who are physically disabled to speak/hear. There are thousands of people who will speak through ISL. In this paper, an attempt is made for them to communicate with other people with the help of Computer Vision (CV) and Artificial Neural Networks (ANN). CV and ANN are the two fascinating and challenging fields in computer science stream. The gestures of ISL are recorded using a video camera; video camera captures the sequence of frames which contains the signs made by the physically disabled people. This image will be processed to get the significance of the sign which is converted into audio. This will aid the physically disabled people (dumb) to communicate without any pen and paper with other people around them who do not understand ISL.**

***Keywords* - Indian Sign Language, Artificial Neural Networks, Computer Vision, Image Processing.**

I. INTRODUCTION

In India there are thousands of people who communicate through sign language. Currently lot of work has been done in American Sign Language (ASL), British Sign Language (BSL) and less work has been done in Indian Sign Language (ISL) till date. The main goal of this translator system is to reduce the communication gap between the physically disabled and rest of the world. This system makes use of Computer Vision (CV) and Artificial Neural Network (ANN) in order to achieve this goal.

Sign Language is a well-structured code gesture; every gesture has meaning assigned to it. Sign Language is the only means of communication for deaf and dumb people. With the advancement of science and technology many techniques have been developed not only to minimize the problem of physically disabled people but also to improvise technology in different fields. Many research works, related to Sign languages have been done as for example the American Sign Language, the British Sign Language, the Japanese Sign Language, and so on. But very few works has been done in Indian Sign Language recognition till date [1].

Finding an experienced and qualified interpreters every time is a very difficult task and also unaffordable. Moreover, people who are not deaf or dumb, never try to learn the sign language for interacting with the deaf people. This becomes a cause of isolation of the deaf/dumb people with the rest of the world. But if the computer can be programmed in such a way that it can translate sign language to audio form, the difference between the normal people and the deaf/dumb community can be minimized. We have proposed a system which is able to recognize the various alphabets of Indian Sign Language through Human-Computer interaction giving more accurate results in less time. It will not only benefit the deaf and dumb people of India but also could be used in various applications in different streams of computer science.

II. LITERATURE SURVEY

Different approaches have been used by different researchers for recognition of various hand gestures which were implemented in different fields. Some of the approaches were vision based approaches, data glove based approaches, soft computing approaches like Support Vector Machines, Fuzzy logic, Genetic Algorithm and others like PCA, Canonical Analysis, etc. The whole approaches could be divided into three broad categories - Hand segmentation approaches, Feature extraction approaches and Gesture recognition approaches. Few of the works have been discussed in this paper.

Many researchers [2-5] used skin filtering technique for segmentation of hand. This technique separates the skin colored pixels from the non-skin colored pixels, thus extracting the hand from the background.

Several different approaches have been used in order to design hand gesture recognition systems. Notable approaches involve using input from special electronic gloves [6] or using input from specially marker gloves, or marked hands [7] [8] [9]. The inconvenience of marker-based systems makes them unattractive for every-day use as human computer interfaces. Other systems are based on the extrapolation of complex representation of hand shapes [10]. This approach involves complex computations and therefore is unattractive for real-time and computational bounded applications.

A marker-free, visual hand recognition system was proposed in [11] [12], where the classification is performed in the curvature space. This approach involves finding the boundary contours of the hand and it is robust in scale, translation and rotation, yet it is extremely demanding in terms of computation. In [13] a multi-system camera is used to pick the centre of gravity of the hand and points with maximal distances from the centre provide the locations of the finger tips, which are then used to obtain a skeleton image, and finally for gesture recognition. In [14] a special camera that supplies depth information was used to identify hand gestures. Other computer vision methods used for hand gesture recognition include specialized mappings architecture [15], principal component analysis [16], Fourier descriptors, neural networks, orientation histograms [17], and particle filters [18].

Unlike previous work, we focus on hand gesture classification using inputs from low-resolution digital cameras and classify the acquired images using features extracted by non-computationally intensive image processing techniques. We classify the extracted feature sets using a multilayer feed forward Neural Networks that was trained in a setup phase and is specifically trained to distinguish between predefined gestures. The system operates in moderately noisy environments and in non-uniform backgrounds and is suitable for real-time.

III. ARTIFICIAL NEURAL NETWORKS

An Artificial Neural Network (ANN), usually called "Neural Network" (NN), is a mathematical model or computational model that is inspired by the structure and/or functional aspects of biological Neural Networks. A Neural Network consists of an interconnected group of artificial neurons, and it processes information using a connectionist approach to computation.

A. Structure of a Single Neuron

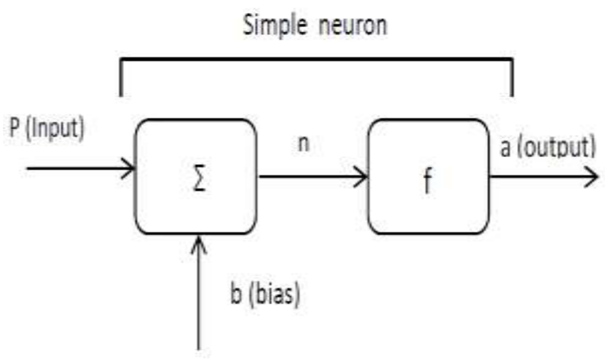

Fig. 1 Structure of a single neuron.

The fundamental building block for neural networks is the single-input Neuron. The following figure 1 shows the structure of such a neuron. There are three distinct functional operations that take place in this example. First, the input *p* is multiplied by the weight *w* to form the product *wp*. Second, the weighted input *wp* is added to the bias *b* to form the net input *n*. Finally, the net input is passed through the transfer function *f*, which produces the output *a*. The names given to these three processes are: the weight function, the net input function and the transfer function.

B. Multilayer feed forward network

The source nodes in the *input layer* of the network supply respective elements of the activation pattern (input vector), which constitute the input signals applied to the neurons (Computation nodes) in the second layer (i.e. the first *hidden layer*). The output signals of the second layer are used as inputs to the third layer, and so on for the rest of the network.

Typically, the neurons in each layer of the network have as their inputs, the output signals of the preceding layer only. The set of output signals of the neurons in the *output layer* of the Network constitutes the overall response of the network to the activation pattern supplied by the source nodes in the input layer.

C. Back propogation Algorithm

The back propagation algorithm is used in layered feed-forward ANN's. This means that the artificial neurons are organized in layers, and send their signals "forward", and then the errors are propagated backwards. The network receives inputs by neurons in the input layer, and the output of the network is given by the neurons on the output layer. There may be one or more intermediate hidden layers. The back propagation algorithm uses supervised learning, which means that we provide the algorithm with examples of the inputs and outputs we want the network to compute, and then the error (difference between actual and expected results) is calculated. The idea of the back propagation algorithm is to reduce this error, until the ANN learns the training data.

The training begins with random weights, and the goal is to adjust them so that the error will be minimal. The explanation here is intended to give an outline of the process involved in Back propagation algorithm. The NN explained here contains three layers. These are input, hidden, and output layer. During the training phase, the training data is fed into to the input layer. The data is propagated to the hidden layer and then to the output layer. This is called the forward pass of the Back propagation algorithm. In forward pass, each node in hidden layer gets input from all the nodes from input layer, which are multiplied with appropriate weights and then summed.

The output of the hidden node is the nonlinear transformation of this resulting sum. Similarly each node in output layer gets input from all the nodes of the hidden layer, which are multiplied with appropriate weights and then summed. The output of this node is the non-linear transformation of the resulting sum. The output values of the output layer are compared with the target output values. The target output values are used to teach network. The error between actual output values and target output values is calculated and propagated back toward hidden layer. This is called the backward pass of the Back propagation algorithm.

The error is used to update the connection strengths between nodes, i.e. weight matrices between input hidden layers and hidden-output layers are updated. During the testing phase, no learning takes place i.e., weight matrices are not changed.

IV. PROPOSED SYSTEM ARCHITECTURE

The block diagram of the proposed system is given in Fig. 2 which comprises of mainly four phases: Video acquisition, Sign extraction, Classification and Recognition and Translation. In our proposed system, we have considered 24 alphabets of Indian sign language, each with 10 samples thus a total of 240 images are used to train the neural network. Some of the samples of signs in Indian Sign Language is shown in Fig. 3 which represents alphabets.

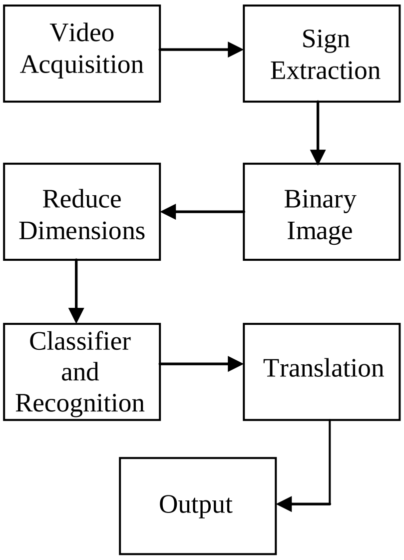

Fig. 2 – Architecture of proposed system.

Fig 3 – ISL alphabets.

A. *Video acquisition*

This is the first phase of the proposed system. The main purpose of this phase is to interact with USB video camera to capture image frames and also to provide user interface to the user through which he can interact with the system. The signs are recorded using a video camera. The image frames at regular interval are recorded from the video camera. The recorded frames are also displayed to the user continuously. These images will contain the signs which need to be processed. These images will be sent to Sign extraction where the images will be segmented.

B. *Sign extraction*

The second phase for our proposed system is the sign extracting stage of the input image which extracts out the skin coloured pixels from the non-skin coloured pixels. This method is very much useful for detection of hand, face etc. The steps carried out for performing skin filtering is given in Fig. 4.

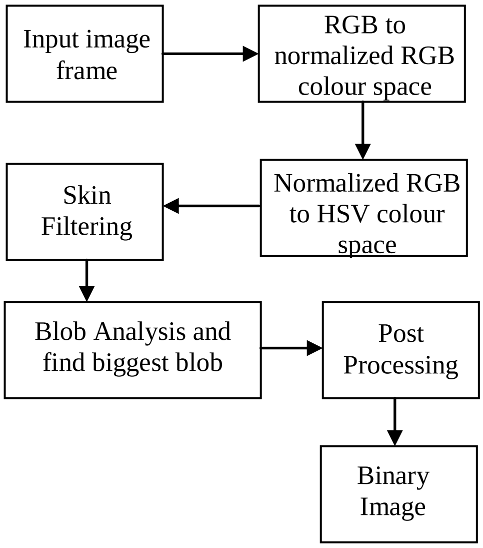

Fig. 4 – Basic block diagram of sign extraction.

Initially, image is converted into normalized RGB. This can be done by calculating the mean of red channel, blue channel and green channel of the image and RGB of individual pixels is divided with this mean and multiplied with a constant value. The resulting value is assigned to the corresponding pixel. This will be done for all pixels. Next, the image frame in normalized RGB representation is transformed into HSV (Hue Saturation Value) representation.

After getting image frame in HSV representation, skin coloured pixels are identified. The value for skin coloured pixels that we have found is, H is greater than 0 and less than 30, S is greater than 80 and less than 195 and V is greater than 40 and less than 255 [measured using openCV library functions]. The pixels in this range are identified as skin. These pixels are assigned white and all other pixels are assigned black. So we have obtained the black and white image by doing this.

Now we will convert HSV image into single channel image. After getting the gray scale image we have to find all the blobs in the image frame. Based on the area of blobs, find biggest blob, draw contours around that region and assign white colour to it and remaining all pixels are set to black colour. So now the image frame only contains hands in it. Later the morphological operations like erosion, dilation and smoothening are applied on to the image frame which will remove the noisy pixels from image frame. To reduce the processing time the image frame needs to be resized into a smaller size that is 100×100 which is sufficient to accurately identify a gesture which will be done in the reduce dimensions stage.

C. *Classifier and recognition*

The neural network that we have used is Multi-Layer Perceptron with three layers, input, hidden and output layer. The input layer consists of 10,000 neurons, the hidden layer contains 12 neurons and the output layer contains 24 neurons.

Since our system is trained to recognize 24 signs of Indian Sign Language, hence 24 neurons are used in the output layer. Each neuron in network has a sigmoid activation function. Back propagation algorithm is used to train the network. A linear transfer function is used by all the neurons in the network. The network is also given training to abort in case of images full of either black or white coloured pixels.

Artificial Neural Network in this phase is designed and trained in Neuroph studio. Neuroph studio will save the trained neural network in a file with extension nnet. The trained neural network is loaded in this phase for recognition. The output of this neural network is the probability of input image matching the images in the data set which are used to train the network. We calculate the highest probability to recognize the sign and also maps the corresponding string associated with matched image in data set.

D. Translation

This is the final phase of the proposed system. Here, open source speech library is used to generate audio from the string. FreeTextToSpeech library is used. It has audio synthesizer with it and also contain pre-loaded audio. After identifying the string of the matched image, it will be sent to this block. This library provides a method called speak which will take string as argument and will play its equivalent audio through the speakers of the computer system.

V. RESULT ANALYSIS

Different image frames were tested and found that the technique of classification was found to show 97% accuracy. Experimentation is carried out in an Intel Core i3 processor machine with 2.27GHz, 3.00GB RAM. The Graphical User Interface (GUI) of the sign language recognition has been shown in the diagram below.

Fig. 4 – GUI of the proposed system.

The some of the results of the proposed system is shown in the below tables.

Table 1 – System Results for sign A

Test Sign	Trained Signs	Matching Probability	Identified Sign
A	**A**	**0.9351310649340291**	"A"
	B	1.287103723580358E-8	
	C	2.560519485869917E-6	
	D	1.4844547347549938E-6	
	E	2.260356567362167E-4	
	F	5.79484677577003E-5	
	G	2.0841853495437452E-12	
	I	3.678867027022147E-5	
	K	9.051317932911418E-7	
	L	1.1058668049458292E-8	
	M	1.8511420720555513E-9	
	N	4.789046718647478E-11	
	O	4.9794985635609284E-8	
	P	3.2568038531242986E-5	
	Q	1.1963807137699307E-12	
	R	1.7656641561647676E-9	
	S	5.927416580645484E-16	
	T	1.9942433444591335E-16	
	U	1.952504952912534E-13	
	V	1.4985264638992737E-14	
	W	6.352524214539187E-8	
	X	8.952993641596612E-9	
	Y	8.305752552826769E-10	
	Z	1.5418896121797374E-14	

Table 2 – System Results for sign L

Test Sign	Trained Signs	Matching Probability	Identified Sign
L	A	2.0154168260111632E-14	"L"
	B	3.840165206353759E-15	
	C	1.0212942236040878E-8	
	D	8.502363109666473E-12	
	E	3.773481192912582E-8	
	F	7.927416953780989E-11	
	G	4.609291473160574E-8	
	I	3.8726664439518476E-7	
	K	3.848877454839964E-7	
	L	**0.9932809679738206**	
	M	6.672358035509669E-14	
	N	3.4425540617334883E-15	
	O	2.3797439548191948E-5	
	P	7.435543706472158E-8	
	Q	4.6041223059136837E-4	
	R	3.7169815014772732E-9	
	S	1.304637692851835E-16	
	T	1.0584853983788372E-8	
	U	7.869083186234963E-8	
	V	1.1880855533001642E-10	
	W	2.3630616669342643E-15	
	X	1.549594495275793E-10	
	Y	2.1291925052525525E-7	
	Z	3.085695258074089E-9	

There are some of the constraints for the system to work properly, some of them are as follows: the palms only be exposed to the camera since we have not used the hand cropping and the background should not have skin coloured pixels which will affect the systems accuracy to extract hands from the background.

VI. CONCLUSION AND FUTURE SCOPE

Sign language is a useful tool to ease the communication between the deaf or mute community and the normal people. Yet there is a communication barrier between these communities with normal people. This project aims to lower the communication gap between the deaf or mute community and the normal world. An Indian Sign Language (ISL) translation system has been implemented in this project. The system makes uses of computer vision, where segmentation will be performed on Hue Saturation Value (HSV) representation of images. A feed forward back propagation Artificial Neural Network (ANN) has been used for the identification of gestures or signs of ISL and will be translated into audio. This system is feasible and will be able to deliver better results than the existing systems in terms of accuracy and efficiency.

The system implemented in this project will only translate the static gestures of ISL from the live video. In the coming days, the dynamic gestures can also be translated by making suitable modifications to the system. Along with dynamic hand gestures, the facial expressions can also be captured using camera as it also contributes to the significance of the signs of Indian Sign Language. The accuracy of the system is acceptable, further a comprehension technique can be used for segmentation of hands from the background to yield 100% accurate results. The position of the hand with the body also contributes to the meaning of the sign. Hence in the future, upper body can also be considered by making considerable amount of changes to our system.

ACKNOWLEDGEMENT

I wish to acknowledge my parents, Mr. Manohar Mullur and Mrs. Mallamma Mullur, for sharing their pearls of wisdom and enlightened me throughout the course of this work. I also wish to thank my friends, who provided insight and expertise that greatly assisted for me to carry out this work.

REFERENCES

[1] Joyeeta Singha and Karen Das, " Indian Sign Language Recognition Using Eigen Value Weighted Euclidean Distance Based Classification Technique," *(IJACSA) International Journal of Advanced Computer Science and Applications*, Vol. 4, No. 2, 2013.

[2] R. Gopalan and B. Dariush, "Towards a Vision Based Hand Gesture Interface for Robotic Grasping", *The IEEE/RSJ International Conference on Intelligent Robots and Systems*, October 11-15, 2009, St. Louis, USA, pp. 1452-1459.

[3] T. Kapuscinski and M. Wysocki, "Hand Gesture Recognition for Man-Machine interaction", *Second Workshop on Robot Motion and Control*, October 18-20, 2001, pp. 91-96.

[4] D. Y. Huang, W. C. Hu, and S. H. Chang, "Vision-based Hand Gesture Recognition Using PCA+Gabor Filters and SVM", IEEE Fifth International Conference on *Intelligent Information Hiding and Multimedia Signal Processing*, 2009, pp. 1-4.

[5] C. Yu, X. Wang, H. Huang, J. Shen, and K. Wu, "Vision-Based Hand Gesture Recognition Using Combinational Features", IEEE Sixth International Conference on *Intelligent Information Hiding and Multimedia Signal Processing*, 2010, pp. 543-546.

[6] *Dataglove model 2 operating manual*. VPL Research Inc., 1989.

[7] V. C. Tartter and K. C. Knowlton. –" Perception of sign language from an array of 27 moving spots|| ", *Nature*, pp (239,676,678), Feb. 19, 1981.

[8] R. Cipolla, Y. Okamoto, and Y. Kuno. – "Robust structure from motion using motion parallax", *In Proc. 4th Intl. Conf. Computer Vision.*

[9] B. Dorner. – "Hand shape identi_cation and tracking for sign language interpretation|| ", *In Looking at people workshop*, Chambery, France, 1993. IJ-CAI.

[10] J. Triesch and C. von der Malsburg. – "A System for Person-Independent Hand Posture Recognition against Complex Background", *IEEE, 2001 Transactions on Pattern Analysis and Machine Intelligence*, Vol.23, No. 12, December 2001.

[11] E. Sanchez Nielsen, L. Antón Canalis, M. Hernandez Tejera, -"Hand gesture recognition for human machine interaction|| ", *Journal of WSCG*, Vol.12, No.1-3 (February 2003).

[12] C. Chang, Y Chen, and Huang, "Hand Pose Recognition Using Curvature Scale Space", *IEEE International Conference on Pattern Recognition*, 2002.

[13] A. Utsumi, T. Miyasato and F. Kishino, "Multi-Camera Hand Pose Recognition System Using Skeleton Image", *IEEE International Workshop on Robot and Human Communication*, pp. 219-224, 1995.

[14] Aviad Barzilai, Adi Fuchs, Michael Kolomenkin, "Hand Gesture Classification Using Depth Information", *CGM Laboratory undergraduate project*, Electrical Engineering Dept., Technion – IIT, 2009

[15] R. Rosales, V. Athitsos, L. Sigal, and S. Sclaroff, "3D Hand Pose Reconstruction Using Specialized Mappings", *IEEE International Con. on Computer Vision*, pp. 378- 385, 2001.

[16] C. Tomasi, S. Petrov, and A. Sastry, "3D Classification + Interpolation", *IEEE International Conf. on Computer Vision*, 2003.

[17] W. T. Freeman and M. Roth, "Orientation Histograms for Hand Gesture Recognition", *IEEE International Conf. on Automatic Face and Gesture Recognition*, 1995.

[18] L. Bretzner, I. Laptev, and T. Lindberg, "Hand Gesture Recognition using Multi-Scale Color Features, Hierarchical Models and Particle Filtering", *IEEE International Conf. on Automatic Face and Gesture Recognition*, 2002.