A Visual Recognition of Static Hand Gestures in Indian Sign Language based on Kohonen Self-Organizing Map Algorithm

Deepika Tewari, Sanjay Kumar Srivastava

***Abstract-** Indian Sign Language (ISL) or Indo-Pakistani Sign Language is possibly the prevalent sign language variety in South Asia used by at least several hundred deaf signers. It is different in the phonetics, grammar and syntax from other country's sign languages. Since ISL got standardized only recently, there is very little research work that has happened in ISL recognition. Considering the challenges in ISL gesture recognition, a novel method for recognition of static signs of Indian sign language alphabets and numerals for Human Computer Interaction (HCI) has been proposed in this thesis work. The developed algorithm for the hand gesture recognition system in ISL formulates a vision-based approach, using the Two-Dimensional Discrete Cosine Transform (2D-DCT) for image compression and the Self-Organizing Map (SOM) or Kohonen Self Organizing Feature Map (SOFM) Neural Network for pattern recognition purpose, simulated in MATLAB. To design an efficient and user friendly hand gesture recognition system, a GUI model has been implemented. The main advantage of this algorithm is its high-speed processing capability and low computational requirements, in terms of both speed and memory utilization.*

***Keywords-** Artificial Neural Network, Hand Gesture Recognition, Human Computer Interaction (HCI), Indian Sign Language (ISL), Kohonen Self Organizing Feature Map (SOFM), Two-Dimensional Discrete Cosine Transform (2D-DCT).*

I. INTRODUCTION

A sign language (also signed language) is a language which, instead of auditorily conveyed sound patterns, uses manual communication and body language to convey meaning. This can include simultaneously combining hand shapes, orientation and movement of the hands, arms or body, and facial expressions to fluidly express a speaker's thoughts [13].A translator is usually needed when an ordinary person wants to communicate with a deaf one. One of the long term goals of gesture recognition is to develop a vision based sign recognition system that can recognise a subset of an existing sign language and translate it to text format.

It is an important research area not only from engineering point of view but also for its impact on the society. Sign languages are non verbal visual languages, different from spoken languages, but they serve the same function. There are different sign languages all over the world such as American Sign.Language (ASL), British Sign Language (BSL), Japanese Sign Language family (Japanese, Taiwanese and Korean Sign Languages), French Sign Language family (French, Italian, Irish, Russian and Dutch Sign Languages), Australian Sign Language, etc. Similarly Indian Sign Language was also developed for Indian deaf community. It is different in the phonetics, grammar, hand gestures and syntax from other country's sign languages. Designing a hand gesture recognition system for ISL is more challenging than other sign languages due to the following reasons [1]:

- ISL uses both hands to make gestures to represent most of the alphabets.
- ISL uses static and dynamic hand gestures.
- Facial expressions are also included.
- One hand moves faster than the other at times in dynamic hand gestures.
- Many of the gestures result in obstruction.
- Complicated hand shapes.
- Locations of the hand with respect to body contribute to the Sign.
- Head/Body postures.
- ISL Involves both global and local hand motion.

Because of these challenges, there is not much research work has been done in ISL recognition system. Considering all the prospects, this paper presents a technique for vision based static hand gesture recognition system in ISL. In this technique, features are extracted from hand gesture images based on skin pixels through image compression using two-dimensional Discrete Cosine Transform. A Kohonen Self Organizing Feature Map (SOFM), an unsupervised learning technique in artificial neural network, is used for classification of DCT-based feature vectors. To make this hand gesture recognition system more user friendly, a GUI model has been designed.

II. RELATED WORK

Many novel and interesting applications of hand gesture recognition have been introduced in recent years. Generally, such systems are divided into two basis approaches namely glove based and vision based approaches. In glove based analysis, detection of the hand is eliminated by the sensors on the hand and 3D model of the hand is subjected to the virtual world so that body motion can be easily captured. Christopher Lee and Yangsheng Xu [16] developed a glove-based gesture recognition system that was able to recognize 14 of the letters from the hand alphabet, learn new gestures and able to update the model of each gesture in the system in online mode, with a rate of 10Hz.

Manuscript published on 30 December 2012.
* Correspondence Author (s)
Deepika Tewari: M.Tech student, Electronics & Communication Engg. Department, Kanpur Institute of Technology, Kanpur, India.
Sanjay Kumar Srivastava, Assistant Professor, Electronics & Communication Engg. Department, Kanpur Institute of Technology, Kanpur, India.

Over the years advanced glove devices have been designed such as the Sayre Glove, Dexterous Hand Master and PowerGlove. The most successful commercially available glove is by far the VPL DataGlove .On the other hand vision-based analysis is more natural and useful for real time applications. Sushmita Mitra and Tinku Acharya had provided a survey on gesture recognition, with detailed description on hand gestures and facial expression [7].

J. Rekha, J. Bhattacharya and S. Majumder proposed an automatic gesture recognition approach for Indian Sign Language [4]. This approach distinguishes the classes (alphabets), by considering local-global configuration of each hand gesture through finger movement and count. Also shape and texture information is used to accurately predict the alphabets signed by the signer. M Geetha and U C Manjusha presented a method for recognition of Indian sign language alphabets and numerals using B-Spline approximation [1]. P.V.V.Kishore, P.Rajesh Kumar, E.Kiran Kumar and S.R.C.Kishore proposed a sign language recognition system for transforming signs of Indian sign language in to voice commands using hand and head gestures of humans [3].Database of extracted features are compared with input video of the signer using a trained fuzzy inference system.

III. SYSTEM OVERVIEW

The objective discussed in this paper is a vision based identification of the static signs of Indian sign language alphabets and numerals. The system deals with images of bare hands which helps the user to interact with the system in a natural way. Indian sign language is composed of static and dynamic hand gestures. A static gesture is observed at the spurt of time. A dynamic gesture is intended to change over a period of time. All the static and dynamic gestures are interpreted over a period of time to understand a full message. This complicated process is termed as hand gesture recognition. Fig. 1 shows block diagram of a Hand Gesture Recognition System.

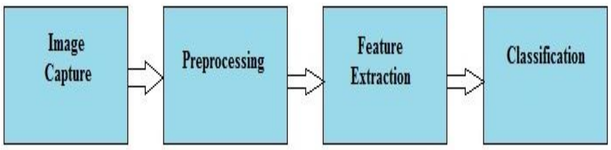

Fig. 1. Block diagram of a Hand Gesture Recognition System

In ISL, there are 26 alphabets in which J and H are dynamic signs. Our algorithm focuses only on static signs for the alphabets and numbers. Since there are no resources to download ISL alphabet image dataset, after a hard effort in getting the dataset various resources, the Indian Sign Language database has been made on our own by capturing images with the help of a digital camera Samsung DV300F. Signs of numerals (0-5) and alphabets (A-Z) excluding J and H sign are shown in fig. 2.

Fig. 2. Indian Sign Language (ISL) Dataset

As these images are not taken in a controlled lightening environment, images are preprocessed in Adobe Photoshop CS3. The algorithm considers hand region as the area of interest. We are assuming that no other objects other than the Region of Interest are present in the background.

Since skin colour in humans varies by individual, research has revealed that intensity rather than chrominance is the main distinguishing characteristic. The recognition stage typically uses an intensity (grayscale) representation of the image compressed by the 2D-DCT for further processing. This grayscale version contains intensity values for skin pixels.

In the first stage, the 2D-DCT for each hand gesture image is computed, and feature vectors are formed from the Discrete Cosine Transform (DCT) coefficients. The second stage uses a self-organizing map (SOM) with an unsupervised learning technique to classify vectors into groups to recognize if the subject in the input image is "present" or "not present" in the image database. If the image is classified as present, the best match image found in the training database is displayed as the result, else the result displays that the image is not found in the image database. Fig. 3 shows schematic view of hand gesture recognition system proposed in this paper.

The proposed hand gesture recognition algorithm has following steps-

- Image Capture
- Preprocessing
- Feature Extraction of the processed image
- Classification

A. Image Capture- Hand gesture images representing signs for different alphabets are taken with a Samsung DV300F 16.0 megapixel digital camera because pictures captured by a webcam are blurred. This digital camera is a CCD camera. The resolution of grabbed image is very high. Images are varied from each other in terms of format, size and resolutions.

Most of the gestures/signs are performed using both hands. Each gesture is performed at various scales, translations, and a rotation in the plane parallel to the image-plane. Since we are assuming that there is no object in the image, other than the hand gesture we need a uniform color background for the ease of segmentation. The signer is required to wear a full sleeve black T-Shirt and a black bandage around the wrist. This provides uniform color background for the experiment.

Published By:
Blue Eyes Intelligence Engineering and Sciences Publication (BEIESP)

The input hand gesture images were then transferred from the digital camera the computer for further processing.

Fig. 4 shows input images representing signs for the alphabets 'M', 'A' and 'G' respectively in Indian Sign Language.

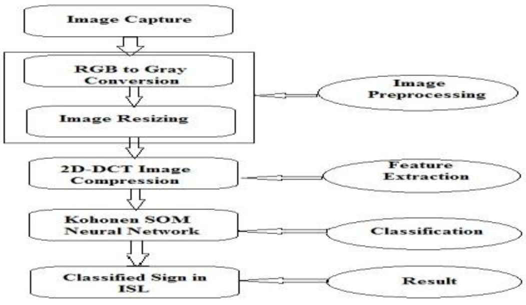

Fig. 3. Schematic View of Proposed Hand Gesture Recognition System for ISL

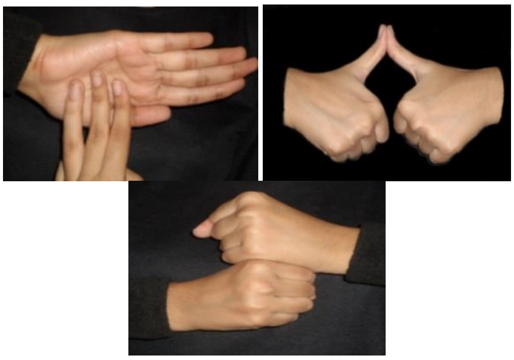

Fig. 4. Input images representing signs for alphabets 'M', 'A' and 'G' in ISL

B. Preprocessing- As these hand gesture images are not taken in a controlled lightening environment, images are preprocessed in Adobe Photoshop CS3. Since images are taken with a digital camera, they have different sizes and different resolutions. So in image preprocessing, brightness and contrast are adjusted using Adobe Photoshop CS3. We convert RGB color images into gray scale images here for ease of segmentation problem. Then the images are resized to 512 × 512 pixels and saved in jpeg format. Fig. 5 shows preprocessed input images of hand gestures.

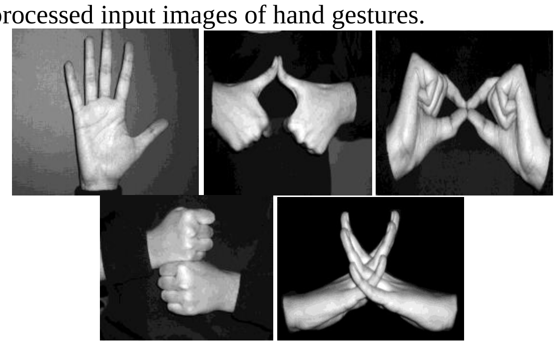

Fig. 5. Preprocessed Input images of Hand Gestures

C. Feature Extraction of the processed image- After preprocessing of images, we can extract the skin regions by either pixel-based or region based extraction method. In this hand gesture recognition system, we have used an intensity (grayscale) representation of the segmented image for further processing. This grayscale version, also called a "skin map," contains intensity values for skin pixels and the background is represented as black. Then, the Two-Dimensional Discrete Cosine Transform (2D-DCT) for each region is computed, and feature vectors are formed from the DCT coefficients. Fig. 6 shows an image and it's DCT compressed image.

Fig. 6. Original Image and it's 2D-DCT Compressed Image (Zoomed)

The DCT is a widely used transformation for data compression. It is an orthogonal transform, which has a fixed set of image independent basis functions, an efficient algorithm for computation, and good energy compaction and correlation reduction properties. Ahmed et al found that the Karhunen Loeve Transform (KLT) basis function of a first order Markov image closely resemble those of the DCT. They become identical because the correlation between the adjacent pixel approaches to one [5].

The DCT can be extended to the transformation of 2D signals or images. This can be achieved in two steps:

- By computing the 1D-DCT of each of the individual rows of the two dimensional image,
- After the above step, by computing the 1D-DCT of each column of the image.

If represents a 2D image of size x (n1, n2) N × N , then the 2D-DCT of an image is given by:

$$Y[z,k] = C[j]\, C[k] \sum_{m=0}^{N-1} \sum_{n=0}^{N-1} x[m,n]\ cos\left(\frac{(2m+1)j\pi}{2N}\right) cos\left(\frac{(2n+1)j\pi}{2N}\right)$$

……………..(1.1)

where j, k, m, n= 0,1,2,..., (N-1) and

$$\text{where } C[j] \text{ and } C[k] = \begin{cases} \sqrt{\frac{1}{N}}, & j,k=0 \\ \sqrt{\frac{2}{N}}, & j,k=1,2,\dots N-1 \end{cases}$$

…………(1.2)

Similarly the 2D-IDCT can be defined as-

$$x[m,n] = \sum_{j=0}^{N-1} \sum_{k=0}^{N-1} C[j]C[k]y[j,k]\ cos\left(\frac{(2m+1)j\pi}{2N}\right) cos\left(\frac{(2n+1)j\pi}{2N}\right)$$

………….(1.3)

The DCT presented in equations (1.1) and (1.3) is orthonormal and perfectly reconstructing provided the coefficients are represented to an infinite precision.

Published By:
Blue Eyes Intelligence Engineering
and Sciences Publication (BEIESP)

This means that when the coefficients are compressed, it is possible to obtain a full range of compressions and image qualities. The coefficients of the DCT are always quantized for high compression, but DCT is very resistant to quantization errors due to the statistics of the coefficients it produces. The coefficients of a DCT are usually linearly quantized by dividing by a predetermined quantization step.

JPEG stands for the Joint Photographic Experts Group, a standards committee that had its origins within the International Standard Organization (ISO).There are following steps in JPEG image compression [6]-

- Original image is divided into blocks of 8 x 8.
- Pixel values of a black and white image range from 0-255 but DCT is designed to work on pixel values ranging from -128 to 127. Therefore each block is modified to work in the range.
- Equation (1.1) is used to calculate DCT matrix.
- DCT is applied to each block by multiplying the modified block with DCT matrix on the left and transpose of DCT matrix on its right.
- Each block is then compressed through quantization.
- Quantized matrix is then entropy encoded.

D. Classification- Pattern classification can be defined as the categorization of input data into identifiable classes by the extraction of significant features or attributes of the data from a background of irrelevant detail. In this paper, we have used self-organizing map (SOM) using an unsupervised learning technique in Artificial Neural Network (ANN) is used to classify DCT-based feature vectors into groups to classify whether the sign mentioned in the input image is "present" or "not present" in the ISL database.

Self-Organizing Map (SOM) is one of the most popular neural network models. It belongs to the category of competitive learning networks. It is based on unsupervised learning, which means that no manual intervention is needed during the learning and that little need to be known about the characteristics of the input data. So we can use the SOM for clustering data without knowing the class memberships of the input data. The SOM can be used to detect features belonging to the problem and thus has also been called SOFM, the Self-Organizing Feature Map (SOFM). We have used the particular kind of SOM known as a **Kohonen Network** in this paper. This SOM has a feed-forward structure with a single computational layer arranged in rows and columns. In the architecture each neuron is fully connected to all the source nodes in the input layer [17]-

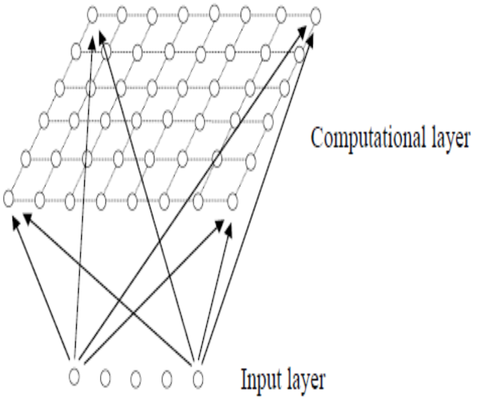

Fig. 7. Kohenon Self-Organizing Map (SOM) [17]

The self-organization process involves four major components [17]-

Initialization: All the connection weights are initialized with small random values.

Competition: In this step for each input pattern, the neurons compute their respective values of a discriminant function which provides the basis for competition. The particular neuron with the smallest value of the discriminant function is declared the winner.

Cooperation: In this step, the active neuron determines the spatial location of a topological neighborhood of excited neurons, thereby providing the basis for cooperation among neighboring neurons.

Adaptation: In last step, the excited neurons decrease their individual values of the discriminant function in relation to the input pattern through suitable adjustment of the associated connection weights, such that the subsequent application of a similar input pattern is enhanced.

For hand gesture recognition process, trained images are reconstructed using weight matrices and recognition is through untrained test images based on minimum Euclidean distance as the similarity measure. Training and testing for this system is performed using the MATLAB Neural Network Toolbox.

IV. EXPERIMENT & RESULTS

Experimentation is carried out in an Intel (R) Core (TM) 2 Duo processor machine with 2.10GHz, 3.00GB RAM, MATLAB R2010a and a 16.0 MP digital camera.

A) Dataset for ISL- A dataset of 35 images is loaded into MATLAB which includes 5 different sign gestures (FIVE, A, B, G and W) with 7 different backgrounds and slightly different postures for the training database. The system is purely data dependent. For testing purpose, we have used some untrained images. Fig. 8 shows image Database used in Proposed Hand Gesture Recognition System in ISL.

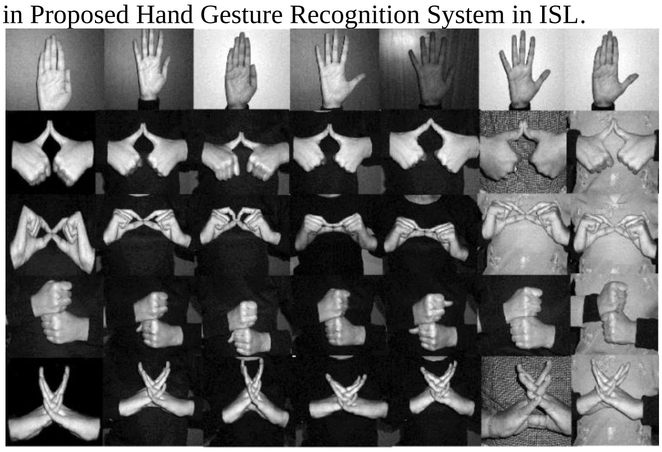

Fig.8. Image Database used in Proposed Hand Gesture Recognition System in ISL

B) *Experiment*- A Graphical User Interface (GUI) has been created to automatically train and recognize the gestures as shown in the fig. 9 below.

Published By:
Blue Eyes Intelligence Engineering
and Sciences Publication (BEIESP)

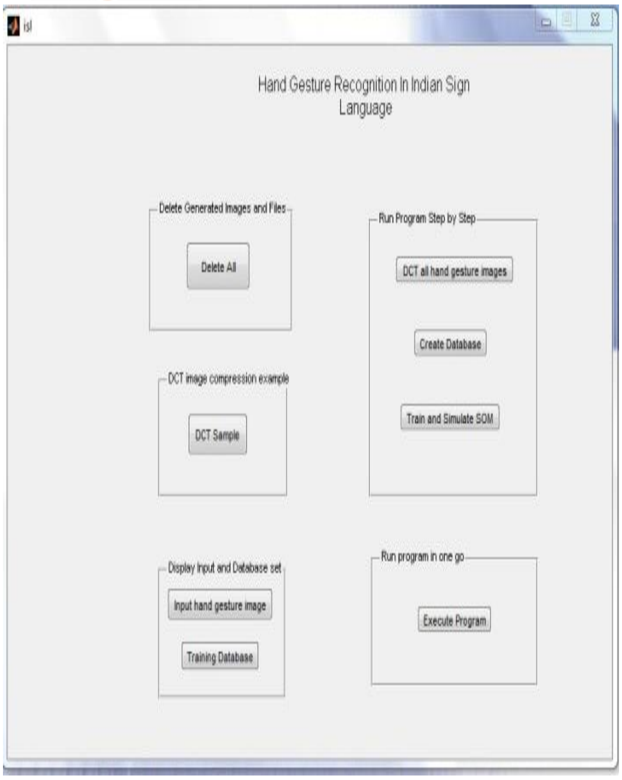

Fig. 9. GUI model of Hand Gesture Recognition System in ISL

The sign language recognition program can be executed by two different ways in the GUI. The first method is through a 3-step push button method in the panel **Run Program Step by Step** and the second method is through a single-step push button method in the panel **Run program in One go**. Both methods will provide the same output result.

At first, all MATLAB workspace variables and all generated image and data files can be deleted from the program directory by pressing the push button labeled **Delete All.** The DCT image compression example is run by pressing the push button labeled **DCT Sample**.

Push button Input hand gesture image shows input test image and Training Database push button displays training database images along with input test image. Fig. 9 shows input test image with the training image database.

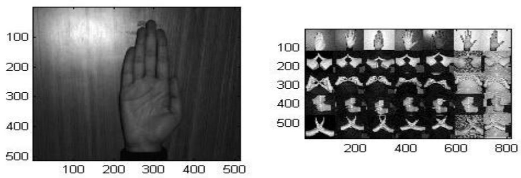

Fig. 10. Input Hand Gesture Image with Image Database

Push button '**DCT all hand gesture images**' executes program to generate DCT images. On pressing push button **'Create Database'** data files are generated in program directory. By pressing push button **'Train and Simulate SOM'**, a single untrained image is given as input into the SOM neural network and the SOM network is simulated for 1000 epochs. The untrained input is matched with the closest image of the same gesture in the training database and generates a correct answer. Fig. 11 and 12 show the SOM layer weights for the 35 hand gesture images in the training database and SOM weight vectors respectively.

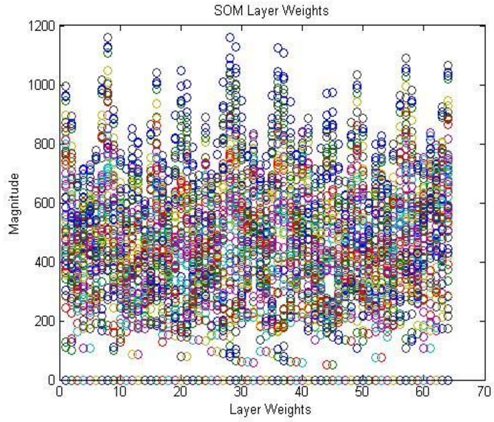

Fig. 11. SOM layer weights for the 35 hand gesture images

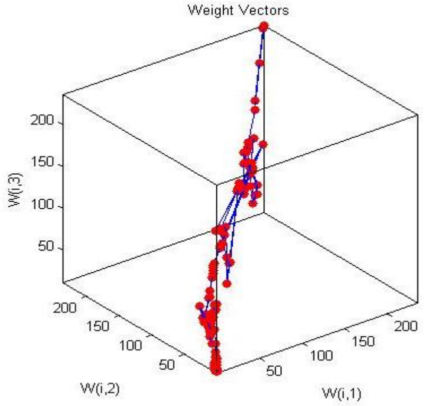

Fig. 12. SOM weight vectors.

Fig. 13. Shows Simulink block model of SOM neural network for the particular input test image.

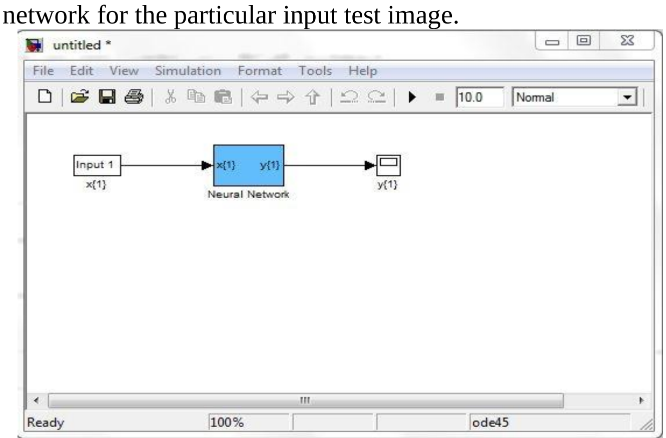

Fig. 13. Simulink block model of SOM neural network

C) Results- Input hand gesture images are changed and the network is simulated repeatedly and the SOM neural network generates correct answers for most of the trials.

Comparison of number of epochs vs. network training time and network accuracy for 10 trials: This experiment is concerned with the processing time of the overall system. Processing time contributes mainly towards the time required for training the SOM network.

Training time depends upon the number of epochs used for training. The aim of this test is to reduce training time, while maintaining the network accuracy rate.

Published By:
Blue Eyes Intelligence Engineering and Sciences Publication (BEIESP)

Table I shows that the best recognition rate achieved with the least amount of processing time is for the case of 1000 training epochs. Recognition rate results obtained are the average of ten consecutive simulations.

Table I. Reducing Processing Time based on Epochs

Number of Epochs	Network Training Time (in seconds)	Recognition rate (for 10 trials) (approx.)
50	5	60%
100	7	60%
500	32	70%
1000	72	80%
1500	123	80%
2000	179	80%

V. CONCLUSION & FUTURE WORK

The Primary focus of this system is to examine image processing as a tool for the conversion of Indian Sign Language (ISL) gesture in to digital text. The proposed gesture recognition system can handle different types of alphabets and number signs in a common vision based platform. The system is suitable for complex ISL static signs. However, it is to be noted that the proposed gesture recognizer cannot be considered as a complete sign language recognizer, as for complete recognition of sign language, information about other body parts i.e., head, arm, facial expression etc is necessary. The experimental results show that the system is sufficient to claim a "working system" for native Indian sign language alphabet & numeral recognition.

In future, there can be many possible improvements that will broaden the scope of this work. First of all, this work can be implemented as a real time application. In improved light conditions and controlled environment, we can develop a real time sign language recognition system. The system developed in this work can be extended too many other research topics in the field of computer vision. We hope this research could trigger more investigations to make computers see and think better.

REFERENCES

[1] Geetha M, Manjusha U C, *A Vision Based Recognition of Indian Sign Language Alphabets and Numerals Using B-Spline Approximation* , IJCSE, Vol. 4 No. 03 March 2012.

[2] Noor A. Ibraheem, Rafiqul Z. Khan, *Vision Based Gesture Recognition Using Neural Networks Approaches: A Review*, International Journal of human Computer Interaction (IJHCI), Volume (3) : Issue (1) : 2012.

[3] J. Rekha, J. Bhattacharya and S. Majumder, *Shape, Texture and Local Movement Hand Gesture Features for Indian Sign Language Recognition* , IEEE 2011.

[5] Bhawna Gautam, *"Image Compression using Discrete Cosine Transform and Discrete Wavelet Transform*" National Institute Of Technology, Rourkela, May 2010.

[6] Swastik Das and Rasmi Ranjan Sethy, *Digital Image Compression Using Discrete Cosine Transform & Discrete Wavelet Transform*, National Institute Of Technology, Rourkela, 2009.

[7] Tinku Acharya, Senior Member, IEEE and Sushmita Mitra, Senior Member, IEEE, "*Gesture Recognition: A Survey*", IEEE Transaction on Systems, Man and Cybernetics- Part C: Applications and Reviews, VOL. 37, NO. 3, May 2007.

[8] Self-Organizing Maps, Wikipedia Source, from http://en. wikipedia.org/wiki/Self-organizing_map (accessed in 2012).

[9] Mathsworks Image Processing Toolbox- DCT and Image compression (accessed in 2012).

[10] Mathworks Image Processing Toolbox.

[11] Mathworks Neural Network Toolbox.

[12] Artificial Neural Network, Wikipedia source from-http://en.wikipedia.org/wiki/Artificial_neural_network (accessed in 2012).

[13] Sign Languages, Wikipedia source from-http://en.wikipedia.org/wiki/Sign_language (accessed in 2012).

[14] Fitzgibbon, A.W. and Lockton, R. *Hand Gesture Recognition Using Computer Vision*, BSc. Graduation Project, Oxford University.

[15] eature extraction Wikipedia source from-http://en.wikipedia.org/wiki/feature_extraction (accessed in 2012).

[16] Christopher Lee and Yangsheng Xu, *Online, interactive learning of gestures for human robot interfaces*, Carnegie Mellon University, The Robotics Institute, Pittsburgh, Pennsylvania, USA, 1996

[17] Bullinaria, John A., Self Organizing Maps: Fundamentals, 2004.

Ms. Deepika Tewari is a student of M.Tech (ECE) in Kanpur Institute of Technology, Kanpur. She is doing her research work in Sign Language Recognition. Her areas of interest are Human Computer Interaction (HCI), Image Processing, and Artificial Intelligence.

Mr. Sanjay Kumar Srivastava is working as Asst. Professor in Electronics and Communication Department in Kanpur Institute of Technology, Kanpur.

Published By:
Blue Eyes Intelligence Engineering
and Sciences Publication (BEIESP)