Artificial neural network based method for Indian sign language recognition

Conference Paper · April 2013

DOI: 10.1109/CICT.2013.6558259

3 authors:

V. Adithya
Central University of Kerala
6 PUBLICATIONS **54** CITATIONS
SEE PROFILE

P.R. Vinod
College of Engineering Chengannur
3 PUBLICATIONS **52** CITATIONS
SEE PROFILE

Usha Gopalakrishnan
Musaliar College of Engineering and Technology
3 PUBLICATIONS **52** CITATIONS
SEE PROFILE

Artificial Neural Network Based Method for Indian Sign Language Recognition

Adithya V.[#1], Vinod P. R.[#2], Usha Gopalakrishnan[*3]

[#]*Department of Computer Engineering*
College of Engineering Chengannur
Kerala, India
[1]adithya@ceconline.edu
[2]vinod@ceconline.edu

[*]*Department of Computer Engineering*
Musaliar College of Engineering and Technology, Pathanamthitta
Kerala, India
[3]writeto_usha@yahoo.com

***Abstract*— Sign Language is a language which uses hand gestures, facial expressions and body movements for communication. A sign language consists of either word level signs or fingerspelling. It is the only communication mean for the deaf-dumb community. But the hearing people never try to learn the sign language. So the deaf people cannot interact with the normal people without a sign language interpreter. This causes the isolation of deaf people in the society. So a system that automatically recognizes the sign language is necessary. The implementation of such a system provides a platform for the interaction of hearing disabled people with the rest of the world without an interpreter. In this paper, we propose a method for the automatic recognition of fingerspelling in Indian sign language. The proposed method uses digital image processing techniques and artificial neural network for recognizing different signs.**

***Keywords*— Indian sign language, Hand segmentation, Distance transform, Projections, Fourier descriptors, Central moments, Artificial neural network.**

I. INTRODUCTION

Sign language is widely used by people who cannot speak and hear or people who can hear but cannot speak. A sign language is composed of various gestures formed by different hand shapes, movements and orientations of hands or body, or facial expressions. These gestures are used by the deaf people to express their thoughts. But the use of these gestures are always limited in the deaf-dumb community, normal people never try to learn the sign language. This causes a big gap in communication between the deaf–dumb people and the normal people. Usually deaf people seek the help of sign language interpreters for translating their thoughts to normal people and vice versa. But this system is very costly and does not work throughout the life period of a deaf person. So a system that automatically recognizes the sign language gestures is necessary. Such a system can minimize the gap between deaf people and normal people in the society.

There are various sign languages across the world. The sign language used at a particular place depends on the culture and spoken language at that place. Indian sign language (ISL) is used by the deaf community in India. It consists of both word level gestures and fingerspelling. Fingerspelling is used to form words with letter by letter coding. Letter by letter signing can be used to express words for which no signs exist, the words for which the signer does not know the gestures or to emphasis or clarify a particular word. So the recognition of fingerspelling has key importance in sign language recognition. The fingerspelling in Indian sign language consists of both static as well as dynamic gestures which are formed by the two hands with arbitrarily complicated shapes. This paper presents a method for the automatic recognition of static gestures in Indian sign language alphabet and numerals. The signs considered for recognition include 26 letters of the English alphabet and the numerals from 0-9. Indian sign language alphabet and numerals are shown in Fig.1 and Fig.2 respectively.

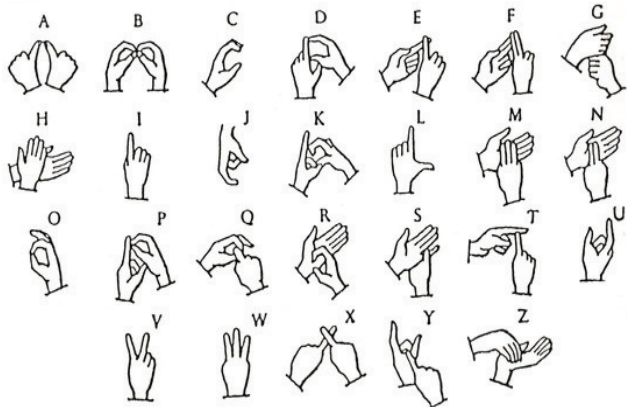

Fig.1 Representation of ISL alphabet

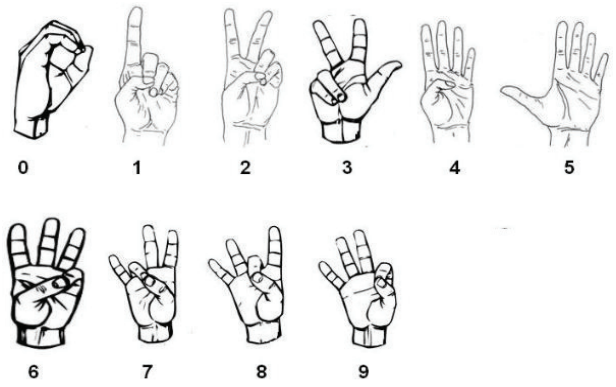

Fig.2 Representation of numerals (0-9)

Sign language recognition systems are broadly classified into two categories: hardware based systems and vision based systems. Hardware based systems require the user to wear some devices to extract features describing the hand sign. Cyber glove is a device which is used for extracting the features such as orientation, movements and colour, of the hands. It is widely used for sign language recognition. Vision based systems use digital image processing techniques to extract features and recognize sign. The method proposed in this paper is a vision based approach in which the user does not have to wear any special device.

This paper consists of five sections: section II discusses about the previous research works done in the area of sign language recognition. The proposed method is presented in detail in section III. The experimental results are given in section IV. Finally section V gives the summary and conclusion of the work.

II. Literature Survey

A glove based system for American Sign Language recognition is proposed by Starner and Pentland [1]. This is considered as the earliest work reported on sign language recognition. They described a system which uses one colour camera to track hands in real time and interprets American Sign Language using hidden markov models. The shape, orientation and trajectory information is used as input to the hidden markov model for the recognition of signed words. The work of Deng-Yuan Huang1 et.al. uses Gabor filters to extract hand gesture features [2]. The principal component analysis is then used to reduce the dimensionality of the feature space. A system for the recognition of Malaysian sign language is proposed by Rini Akmelia, Melanie Po-Leen Ooi and Ye Chow Kuang is given in [3].

A real time Arabic sign language recognition system proposed by Nadia R. Albelwi and Yasser M. Alginahi is presented in [4]. This system uses a Haar like method to track hand in the video frames and the bounded hand region becomes the area of interest. The resultant images are transformed into the spectral domain using Fourier transformation to form the feature vectors and the classification is done using KNN method. In [5] Stephan Liwicki and Mark Everingham presented a method of recognizing the British sign language, in which the hand shapes are described by a modification of the Histogram of Oriented Gradient (HOG) descriptor, as a joint histogram over quantized gradient orientation and position and classification was done using HMM. Azadeh Kiani Sarkalehl et.al. proposed a system for Persian sign language recognition in [6]. This method uses discrete wavelet transform to extract features from the gesture images and a multilayered perceptron neural network for the classification of gestures.

Al-Jarrah and Halawani [7] presented a method for automatic translation of gestures of the manual alphabet in the Arabic sign language. This system utilizes the feature values that comprise of some length measures which indicate the positions of the fingertips. Classification is done using a subtractive clustering algorithm and fuzzy inference system. An approach for the recognition of static alphabetic signs of Spanish sign language is addressed in [8]. Rokade et.al.[9] used thinning method for the recognition of American Sign Language numbers. The works proposed for the recognition of American, Japanese and Korean sign language alphabets are given in [10] [11] and [12] respectively.

A video based Indian sign language recognition system was developed by P. V. Kishore and Rajesh Kumar using wavelet transform and fuzzy logic [13]. They used wavelet based video segmentation technique to detect shapes of various hand signs. Shape features of hand gestures are extracted using elliptical Fourier descriptions and PCA is used to reduce the dimensionality of the feature space. Recognition of gestures from the extracted features is done using a fuzzy inference system. J. Rekha et.al. used a feature descriptor which is a combination of shape, texture and local movements of hand features for the recognition of Indian sign language alphabet. The shape, texture and finger features of each hand are extracted using Principle Curvature Based Region (PCBR) detector. Wavelet Packet Decomposition (WPD-2) and complexity defects algorithms respectively for hand posture recognition process. To classify each hand posture, multi class non linear support vector machine (SVM) was used [14].

Geetha M. and Manjusha U. C. proposed a method [15] for the recognition of Indian sign language alphabet and numerals using B-Spline approximation. Their algorithm approximates the boundary extracted from the region of interest, to a B-Spline curve by taking the maximum curvature Points as the control points. Then the B-Spline curve is subjected to iterations for smoothening resulting in the extraction of key maximum curvature points, which are the key contributors of the gesture shape. Support vector machine (SVM) is used as the classification tool in this method.

III. Proposed Method

In this paper, we propose a method that translates the fingerspelling in Indian sign language to textual form. The proposed method has four major steps.

A. Image Acquisition and Pre-processing
B. Hand Segmentation
C. Feature Extraction
D. Classification

A. Image Acquisition and Pre-processing

Image acquisition is the process of capturing the hand gesture images representing different signs. In this step, the image database is created for training and testing the system. The image dataset of Indian sign language alphabet and

numerals are not available from any resources. So the dataset is made in our lab with suitable lighting and environmental setup. Images are captured with a black background.

Image capturing is a random process. The resolution of various image capturing devices may not be the same. This results in different resolution of the captured images. For accurate comparison of the features and to reduce the computational effort needed for processing, all the images should be scaled to a uniform size.

B. Hand Segmentation

Hand segmentation is the process of extracting the hand sign from the captured image. Efficient hand segmentation has a key role in sign language recognition task. The hand region can be extracted from the background using skin colour based segmentation. Colour based segmentation is computationally simple and the colour descriptor of an object is invariant to geometric transformations such as translation rotation and scaling. So colour is widely used as a powerful descriptor for object detection.

Colour models represent a colour in a standard way. Different colour models and colour based systems have been used for skin detection applications. The proposed method for hand detection is applied in the YCbCr colour space. Inorder to detect the skin colour in the input image it is first converted to YCbCr colour space. YCbCr separates RGB into luminance and chrominance components where Y is the luminance component and Cb, Cr are the chrominance components. RGB values can be transformed to YCbCr colour space using the following equations

$$Y = 0.299R + 0.587G + 0.114B,$$
$$Cr = 128 + 0.5R - 0.418G - 0.081B,$$
$$Cb = 128 - 0.168R - 0.331G + 0.5B.$$

Skin coloured pixels in the input image are identified by applying a thresholding technique based on the skin colour distribution in YCbCr colour space. The colour of each pixel in the image is set to black or white according to the values of Y, Cb and Cr components. If the Y, Cb and Cr values of a pixel are within a predefined range of skin colour, set that pixel as white otherwise black. Thus a pixel is classified as belonging to skin if it satisfies the following relation:

$$75 < Cb < 135 \text{ and } 130 < Cr < 180 \text{ and } Y > 80.$$

The result of segmentation produces a binary image with the skin pixels in white colour and background in black colour. The resulting binary image may contain noise and segmentation errors. Filtering and morphological operations are performed on the input image to decrease noise and segmentation errors if any.

The orientation of the object in the captured images is not always the same. In order to ensure the reliability and to enhance the robustness of gesture recognition, the images must be subjected to coordinate adjustments. For this, the object's major axis should be made parallel to the X-axis of the coordinate system.

C. Feature Extraction

After image segmentation, we get a binary image containing the handshape representing a particular sign. Inorder to classify this image, we need to extract certain features of that image. Shape is an important visual feature of an object. Many different methods are available for the representation and description of a particular shape. In this work we propose a new feature for shape representation. The proposed shape feature is derived from the distance transform of the binary image.

1) Distance Transformation: Distance transform is a derived representation of an image which is normally applied to binary images. It is also known as distance map or distance field. To apply distance transform on an image, it should be first converted to binary form. A binary image contains object pixels and non object pixels. Distance transform of a binary image gives another image of the same size where each pixel value is replaced by the minimum distance of that pixel from its nearest background pixel. So the distance transform of a binary image gives a grayscale image where the gray scale intensity of the foreground region corresponds to the distance from the closest boundary pixel.

The three different distance measures used for finding the distance transform of an image are Euclidean, city block, and chessboard. The Euclidean distance transformation is invariant to rotation of the image. So it is the most commonly used measure for finding the distance transformation, but it involves time consuming calculations such as square, square root and the minimum over a set of floating point numbers. There are many techniques to obtain Euclidean distance transform of an image. Most of these methods are either inefficient or complex to implement. Normally the Euclidean distance transform is computed on the basis of a mathematical morphological approach using gray scale erosions with successive small distance structuring elements by decomposition. The squared Euclidean distance transform is calculated by using a squared Euclidean distance structuring element. The Euclidean distance transform is obtained by applying a square root operation over the squared Euclidean distance transform matrix.

The distance transform of the image is widely used for object feature extraction and recognition tasks. In the proposed method, the distance transform is computed by using the Euclidean distance. The equations for computing the distance between two pixels with coordinates (x, y) and (u, v) are shown below:

The city-block distance between two points P = (x, y) and Q = (u, v) is defined as

$$d_4(P, Q) = |x - u| + |y - v|.$$

The chessboard distance between P and Q is defined as

$$d_8(P, Q) = \max(|x - u|, |y - v|).$$

The Euclidean distance between P and Q is defined as

$$d_e(P, Q) = \sqrt{(x - u)^2 + (y - v)^2}.$$

2) Projections of Distance Transform Coefficients: This step calculates the row projection vector and the column projection vector from the distance transform image. The calculation of projection vectors works as follows:

- Find the sum of the pixel values in the rows and columns of the distance transformed image. This step returns two vectors.

- Row vector R, where each element in this vector is the sum of non-zero pixel values of the corresponding row of the distance transformed image.
- Column vector C, where each element in this vector is the sum of non-zero pixel values of the corresponding column of the distance transformed image.

The resulting row projection vector and column projection vector are the two 1-D functions which uniquely represent the hand shape in the input image. So these two vectors can be considered as shape descriptors. These shape descriptors represent the shape locally and they are sensitive to noise. So these descriptors have to be processed further to make them robust.

3) Fourier Descriptors: Fourier transform coefficients of the shape descriptors form the Fourier descriptors of the shape. In the proposed method, Fourier descriptors of the row and column projection vectors are calculated. These descriptors represent the handshape in the frequency domain. Fourier descriptors have strong discrimination power and also, they overcome the noise sensitivity present in the shape representation. Moreover, Fourier descriptors are information preserving and they can be normalized easily.

For the two vectors R (t) and C (t), where t=0, 1, 2... N-1 the discrete Fourier transforms are given by

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

and N is the size of R.

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

and N is the size of C.

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

4) Feature Vector: The feature values are formed from the Fourier descriptors of the row and column projection vectors by taking only the magnitude of the Fourier coefficients and ignoring the phase information. The feature values are normalized by dividing the magnitude values of all the Fourier coefficients by the magnitude value of the first coefficient which is called the dc component. Although the number of coefficients generated from the transform is usually large, a set of parameters describing the distribution of these coefficients are enough to capture the overall features of the shape. The feature vector for each gesture is formed of six feature values which are the second, third and fourth central moments of the normalized Fourier coefficients of the row and column projection vectors.

Central Moments: Central moments are a set of values which characterize the properties of a probability distribution. The higher order central moments are only related to the spread and shape of the probability distribution, rather than to its location. So they are preferred to ordinary moments for describing the probability distribution. For a real-valued random variable X, the kth moment about the mean or k^{th} central moment is given by $\mu_k = E\left[(XE[X])^k\right]$ where E is the expectation operation. The first few central moments have intuitive interpretations: The zeroth central moment μ_0 is one. The first central moment μ_1 is zero. The second central moment μ_2 is called the variance, and is usually denoted as σ^2, where σ represents the standard deviation of the distribution. The third central moment μ_3 and fourth central moment μ_4 are used to define the standardized moments skewness and kurtosis respectively.

Variance: Variance is a measure of the dispersion of the data in a sample. It is a good descriptor of the probability distribution of a random variable. It describes the spread of the numbers from the mean value. In particular, variance is the second order moment of a distribution. Thus, it can be used as a parameter for distinguishing between probability distributions. Although there are many methods available for representing different distributuions, the moment based methods are preferred due to their computational simplicity. The variance of a random variable or distribution is defined as the expectation, or mean, of the squared deviation of that variable from its expected value or mean.

In general, the variance of a set of data values with finite size N is given by

$$\mu_2 = \frac{1}{N}\sum_{i=1}^{N}(x_i - \mu)^2,$$

where x_i is the sample values, i=1, 2...N.

Skewness: Skewness is a measure of the asymmetry of the probability distribution of a real-valued random variable. The skewness value can have positive or negative values, or it can be undefined also. If the tail on the left side of the probability density function is longer than the right side, it indicates a negative skew. In this case, the bulk of the values possibly including the median lie to the right of the mean. A positive skew results when the tail on the right side is longer than the left side and the bulk of the values lie to the left of the mean. When the values are evenly distributed on both sides of the mean, the skewness becomes zero. This does not always imply a symmetric distribution

The skewness of a set of data values with finite size N is given by

$$\mu_3 = \frac{1}{N}\sum_{i=1}^{N}(x_i - \mu)^3,$$

where x_i is the sample values, i=1, 2...N.

Kurtosis: Kurtosis is a measure of the "peakedness" of a probability distribution. It is a characteristic of the distribution of a real-valued random variable. Similar to the concept of skewness, kurtosis is also a descriptor of the shape of a probability distribution and, there are different methods for quantifying it for a theoretical distribution and corresponding ways of estimating it from a sample from a population.

The kurtosis of a set of data values with finite size N is given by

$$\mu_4 = \frac{1}{N}\sum_{i=1}^{N}(x_i - \mu)^4,$$

where x_i is the sample values, i=1, 2...N.

D. Classification

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

1) *Artificial Neural Network*: An artificial neural network is a computational model inspired by the neural structure of human brain. The processing elements in an artificial neural network are artificial neurons which mimic the biological neurons. In biological neurons, inputs are received through synapses on the membrane of the neuron. When this input signal exceeds a predefined threshold value the neuron is activated and it emits a signal through the axon. This signal is sent to another synapse to activate other neurons. The same principle is employed in the working of artificial neural network.

An artificial neural network processes information by creating connections between artificial neurons. Artificial neural networks have wide application in the area of pattern recognition and they are widely used to model complex relationship between inputs and outputs. Training or learning is used to configure a neural network such that the application of a set of inputs produces a set of desired outputs. Many different algorithms exist to train an artificial neural network. Training strategy can be either supervised or unsupervised. In supervised learning the network is trained using a set of labelled training examples. Unsupervised learning is used to find hidden structure in unlabelled data.

A feed forward neural network in combination with a supervised learning scenario is used in the proposed method. The network has one input layer, one output layer and two hidden layers with each layer fully connected to the following layer. The most commonly used algorithm for training a feed forward neural network is the backpropagation algorithm. It works by the principle of "backward propagation of errors". Backpropagation is a supervised learning technique and the network is provided with the pairs of inputs and outputs that the network has to compute. The input patterns are given to the network through the neurons in the input layer and the output of the network is obtained through the neurons in the output layer. Then the backpropagation algorithm computes the difference between actual and expected results and this error value is propagated backwards. The backpropagation algorithm tries to minimize this error until the neural network learns the training data.

2) *Training Phase:* In our proposed approach, the neural network is trained to classify 36 hand signs. The training dataset contains 360 images with 10 images of each of the 36 signs.

3) *Testing Phase:* In this phase, a dataset containing 180 images corresponding to 5 images of 36 signs each is used to test the proposed system.

IV. Experimental Results

This section gives the implementation results of the proposed sign language recognition system. The system was implemented using MATLABR2010a in a machine with Intel i3, 2.2GHz processor and 4GB RAM. The experiments are conducted for 36 signs with 15 images of each. 10 images of each sign are used for training the system and 5 images of each sign are used for testing the system. Features extracted from the training image set are used for training the feed forward neural network. The distributions of feature values for the 36 signs are shown in the graphs given in Fig.3, Fig.4 and Fig.5.

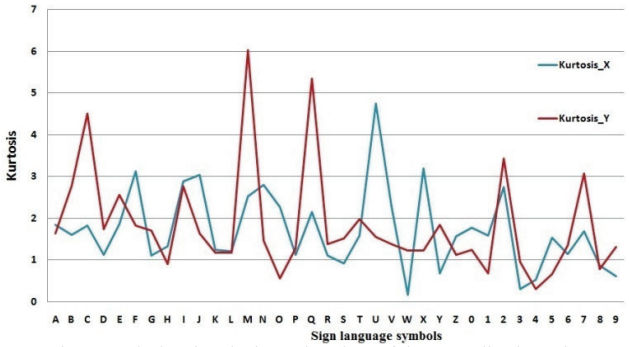

Fig.3 Graph showing the kurtosis values of the normalized Fourier coefficients of the two projection vectors

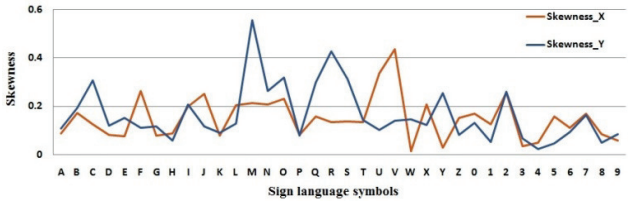

Fig.4 Graph showing the skewness values of the letters normalized Fourier coefficients of the two projection vectors

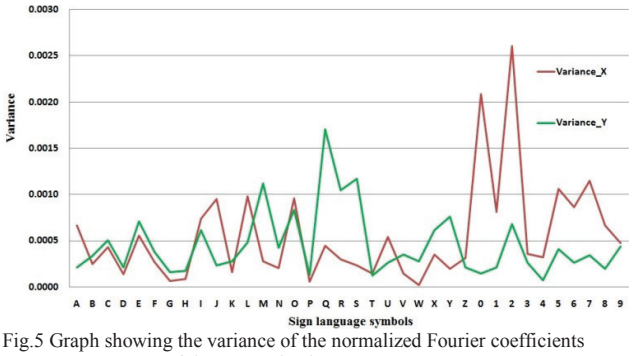

Fig.5 Graph showing the variance of the normalized Fourier coefficients of the two projection vectors

The system is tested using the feature values from the test image set. The recognition rate of the system is estimated as the ratio of the number of signs identified correctly to the total number of signs in the test dataset. We have got an average recognition rate of 91.11% from our experiment. This indicates a good recognition result when considering the large variety of signs in the dataset. The results are depicted in TABLE 1.

TABLE 1
PERFORMANCE OF THE PROPOSED METHOD

Number of test images	Correctly classified	Wrongly classified	Accuracy (%)
36x5 = 180	164	16	91.11

V. CONCLUSION

A neural network based method for automatically recognizing the fingerspelling in Indian sign language is presented in this paper. The signs are identified by the features extracted from the hand shapes. We used skin colour based segmentation for extracting the hand region from the image. A new shape feature based on the distance transform of the image is proposed in this work. The features extracted from the sign image are used to train a feed forward neural network that recognizes the sign. The method is implemented completely by utilizing digital image processing techniques so the user does not have to wear any special hardware device to get the features of the hand shape. Our proposed method has low computational complexity and very high accuracy when compared to the existing methods.

REFERENCES

[1] T. Starner and A. Pentland, "Real-time American sign language recognition from video using hidden markov models", Technical Report, M.I.T Media Laboratory Perceptual Computing Section, Technical Report No. 375, 1995.

[2] Deng-Yuan Huang1, Wu-Chih Hu2, Sung-Hsiang Chang , "Vision-based Hand Gesture Recognition Using PCA + Gabor Filters and SVM" ,2009 Fifth International Conference on Intelligent Information Hiding and Multimedia Signal Processing.

[3] Rini Akmeliawati, Melanie Po-Leen Ooi and Ye Chow Kuang, "Real-Time Malaysian Sign Language Translation Using Colour Segmentation and Neural Network", IEEE on Instrumentation and Measurement Technology Conference Proceeding, Warsaw, Poland 2006, pp. 1-6.

[4] Nadia R. Albelwi, Yasser M. Alginahi, "Real-Time Arabic Sign Language (ArSL) Recognition" International Conference on Communications and Information Technology 2012.

[5] Stephan Liwicki, Mark Everingham, "Automatic Recognition of Fingerspelled Words in British Sign Language", School of Computing University of Leeds

[6] Azadeh Kiani Sarkalehl, Fereshteh Poorahangaryan, Bahman Zan, Ali Karami, "A Neural Network Based System for Persian Sign Language Recognition" IEEE International Conference on Signal and Image Processing Applications 2009.

[7] Al-Jarrah, A. Halawani, "Recognition of gestures in Arabic sign language using neuro-fuzzy systems," The Journal of Artificial Intelligence 133 (2001) 117–138.

[8] Incertis, J. Bermejo, and E. Casanova, "Hand Gesture Recognition for Deaf People Interfacing," The 18th International Conference on Pattern Recognition, 2006 IEEE.

[9] R. Rokade, D. Doye, and M. Kokare, "Hand Gesture Recognition by Thinning Method" International Conference on Digital Image Processing (2009).

[10] R. Feris, M. Turk, R. Raskar, K. Tan, and G. Ohashi. "Exploiting depth discontinuities for vision-based fingerspelling recognition". In IEEE Workshop on Real-time Vision for Human-Computer Interaction, 2004.

[11] M. Lamari, M. Bhuiyan, and A. Iwata. "Hand alphabet recognition using morphological pca and neural networks". In Proc. IJCNN, pages 28392844, 1999.

[12] A. Park, S. Yun, J. Kim, S. Min, and K. Jung. "Real time vision-based Korean finger spelling recognition system". Proc. World Academy of SET, 34:293298, 2008

[13] P. V. V. Kishore and P. Rajesh Kumar, "A Video Based Indian Sign Language Recognition System (INSLR) Using Wavelet Transform and Fuzzy Logic", IACSIT International Journal of Engineering and Technology, Vol. 4, No. 5, October 2012

[14] J. Rekha, J. Bhattacharya, S. Majumder, "Improved Hand Tracking and Isolation from Face by ICondensation Multi Clue Algorithm for continuous Indian Sign Language Recognition", Surface Robotics Laboratory, Central Mechanical Engineering Research Institute, CSIR.

[15] 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), March 2012.