INTERNATIONAL JOURNAL OF SCIENTIFIC RESEARCH IN ENGINEERING AND MANAGEMENT (IJSREM)
VOLUME: 03 ISSUE: 09 | SEPT -2019 ISSN: 2590-1892

Automatic Indian Sign Language Recognition of Finger Spelling Using Artificial Neural Network

[1]Mrs. Nisha G R,Assistant Professor
Department of Electronics and Communication
Vivekananda College of Engineering and Technology, Puttur, 574203

[2]Mrs. Nirupama K,Assistant Professor
Department of Electronics and Communication
Vivekananda College of Engineering and Technology, Puttur, 574203

[3]Mr. Udaya Kumar,
Associate Professor
Department of Electronics and Communication
SIT,
Mangalore

***Abstract*—Indian Sign Language (ISL) 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. It is the only communication mean for the deaf-dumb community people. But the hearing people never try to learn the sign language. This causes the isolation of deaf and dumb 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, Kurtosis, Artificial neural network

I. INTRODUCTION

Sign Language is a well-structured code gesture and is composed of various gestures, every gesture assigned with its own meaning. 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 deaf and dumb people but also to implement it in different fields. As per Joyeeta Singha and Karen Das, many research works related to Sign languages have been done [1] [8]. But till date very few works done in Indian Sign Language (ISL) Recognition. Finding an experienced and qualified interpreter every time is a very difficult task and also unaffordable. If the computer can be programmed in such a way that it can automatically recognizes the sign language, the implementation of such a system provides a platform for the interaction of hearing disabled people with the normal people without an interpreter. Such a system minimizes the gap between normal people and hearing disabled people.

The Key importance in Sign Language is the recognition of finger spelling.The fingerspelling in Indian sign language consists of both static and dynamic gestures which are formed by the two hands with arbitrarily complicated shapes.The signs considered for recognition include 26 letters of the English alphabet and the numerals from 0-9 [2].This paper shows a method to recognize static gestures automatically. Indian sign language alphabet and numerals are shown in Fig.1 and Fig.2 respectively.

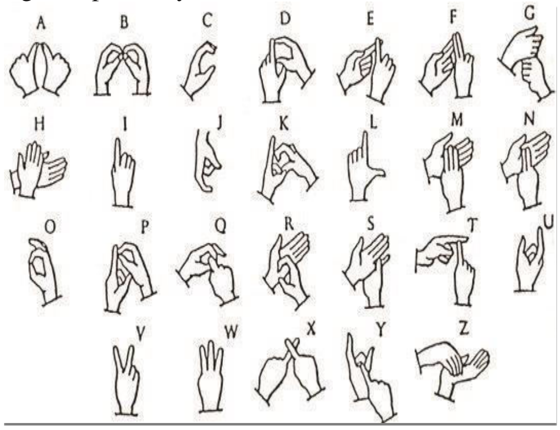

Fig-1: Representation of ISL alphabet

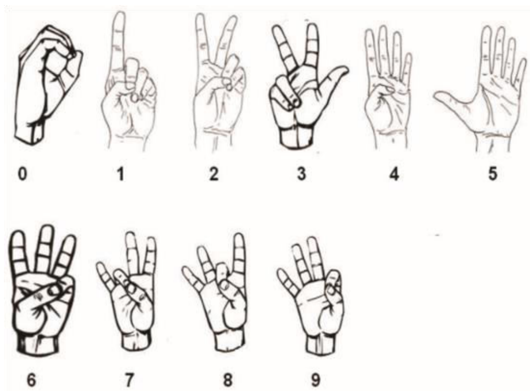

Fig-2: Representation of ISL numerals

ISSN: 2590-1892

II. LITERATURE SERVEY

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 Artificial Neural Network, 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. Indian Sign Language Recognition Using Eigen Value Weighted Euclidean Distance Based Classification Techniqueproposed byJoyeeta Singha, Karen Das.Few of the works have been discussed in this paper. In this paper we outline Hand segmentation and gesture recognition concepts [1]

III PROPOSED METHOD

The objective discussed here is a vision based identification of the static signs of Indian sign language alphabets and numerals. To interact the user with system in a natural way the system must deals with images of bare hands. 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.
The proposed method here translates the finger spelling in Indian sign language to textual form. TheProposed method has shown in below Fig.3

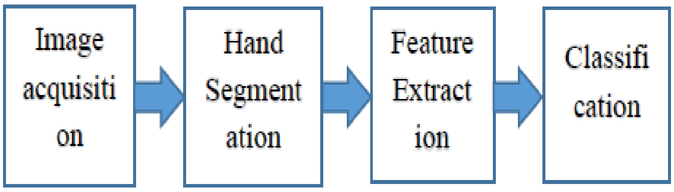

Fig-3: Block diagram of proposed method

3.1. Image Acquisition

The process of capturing hand gestures representing different signs is the term Image acquisition. Hand gesture images representing signs for different alphabets are taken with a webcam. The resolution of grabbed image is very high. Images are varied from each other in terms of format, size and resolutions. We are using both the hands for different signs. 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 colour background for the ease of segmentation [6].Also 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. With suitable lighting and environmental setup the dataset is made in our lab. Images are captured with a black background.

3.2 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 [2] [8]. 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 (1), (2) and (3)

$$Y = 0.299R + 0.587G + 0.114B,\text{-----------------}(1)$$
$$Cr = 128 + 0.5R - 0.418G - 0.081B,\text{-------------}(2)$$
$$Cb = 128 - 0.168R - 0.331G + 0.5B.\text{-------------}(3)$$

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. The final output image of hand segmentation will be Binary image.

3.3 Feature Extraction

Feature Extraction stage is necessary because certain features has to be extracted so that they are unique for each gesture or sign. After the image/Hand segmentation process we get binary image containing the hand shape representing a particular sign. In order to classify this binary image we need to extract some features like shape, movements, orientation etc. Shape is an important feature of any object [6]. So there are many methods are available to describe and represent a particular shape. In this work we propose a shape feature which is derived from the distance transform of the binary image.

3.3.1 Distance Transformation:

Distance transform is abderived 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 [2].

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 in equation (4)

d4(P,Q) = | x - u | + | y - v |----------------------------(4)

The chessboard distance between P and Q is definedas in equation (5)

d8(P,Q) = max(| x - u |,| y - v |)------------------------(5)

The Euclidean distance between P and Q is definedas in equation (6)

de(P,Q) =sqrt((x - u)2 + (y - v)2)-------------------(6)

3.3.2 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.

3.3.3 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 hand shape 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 equation (7) and (8)

$u_n = \frac{1}{N}\sum_{t=0}^{N-1} R(t)exp(-j2\pi nt|N)\,, where\ n = 0, 1, 2 \ldots. N - 1$-----------------(7)

$v_n = \frac{1}{N}\sum_{t=0}^{N-1} C(t)exp(-j2\pi nt|N)\,, where\ n = 0, 1, 2 \ldots. N - 1$-----------------(8)

and N is the size of C.

The coefficients *un* and *vn* are called the Fourier descriptors of the shape.

3.3.4 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 kth central moment is given by equation (9)

$$\mu_k = E[(XE[x])^k] \text{-----------}(9)$$

where E is the expectation operation. The first few central moments have intuitive interpretations: The zeroth central moment μ0 is one. The first central moment μ1 is zero. The second central moment μ2 is called the variance, and is usually denoted as σ^2, where σ represents the standard deviation of the distribution. The third central moment μ3 and fourth central moment μ4 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 equation (10)

$$\mu_2 = \frac{1}{N}\sum_{i=1}^{N}(x_i - \mu)^2 \text{-----------}(10)$$

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 equation (11)

$$\mu_3 = \frac{1}{N}\sum_{i=1}^{N}(x_i - \mu)^3 \text{-------------}(11)$$

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 equation (12)

$$\mu_4 = \frac{1}{N}\sum_{i=1}^{N}(x_i - \mu)^4 \text{---------------}(12)$$

Where $x_i$is the sample values, i=1, 2, ……N

3.4 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.

3.4.1 Artificial Neural Network: An artificial neural network is a computational model inspired by the neural structure of human brain [7]. 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.

3.4.2 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.

3.4.3 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 MATLABR2012a in a machine with Intel i3, 2.2GHz processor and 4GB RAM. The experiments are conducted for 5 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.

Feature extracted from the training image set are used for training the feed forward neural network. The distributions of feature values for the 5 signs are shown in the graphs given in Fig.4, Fig.5 and Fig.6.

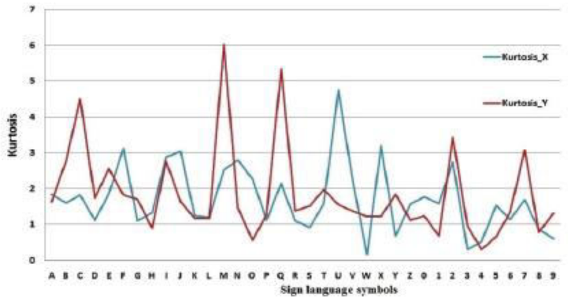

Fig-4: Graph showing kurtosis values of the normalised Fourier coefficient of the two projection vectors

Fig-5: Graph showing Skewness values of the normalised Fourier coefficient of the two projection vectors

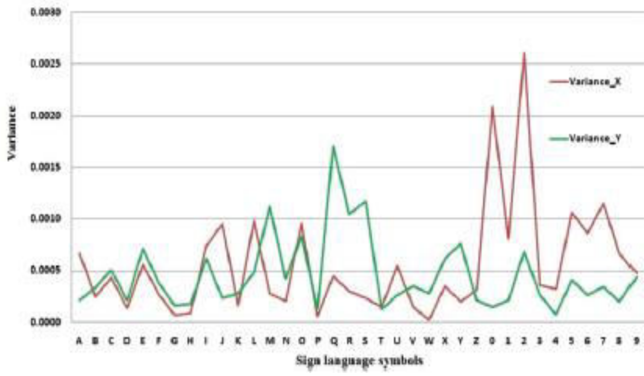

Fig-6: Graph showing Variance values of the normalised Fourier coefficient of the two projection vectors

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 86.66% from our experiment. This indicates a good recognition result when considering the large variety of signs in the dataset.

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] Joyeeta Singha, Karen Das , Indian Sign Language Recognition Using Eigen Value Weighted Euclidean Distance Based Classification Technique, IJACSA, Vol. 4 No. 02 2013

[2] Adithya V, Vinod P. R, Usha Gopalakrishnan, Artificial Neural Network Based Method for Indian Sign Language Recognition, ICT , 2013,

[3] 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.

[4] Padmavathi. S, Saipreethy.M.S, Valliammai.V, Indian Sign Language Character Recognition using Neural Networks, IJCA March 2013.

[5] Archana S. Ghotkar & Gajanan K. Kharate, Hand Segmentation Techniques to Hand Gesture Recognition for Natural Human Computer Interaction, IJHCI, Volume (3): Issue (1): 2012

[6] Deepika Tewari, Sanjay Kumar Srivastava , A Visual Recognition of Static Hand Gestures in Indian Sign Language based on Kohonen Self- Organizing Map Algorithm , IJEAT I , Volume-2, Issue-2, December 2012

[7] Noor A. Ibraheem & Rafiqul Z. Khan, Vision Based Gesture Recognition Using Neural Networks Approaches: A Review, (IJHCI)), Volume (3) : Issue (1) : 2012

[8] Joyeeta Singha, Karen Das, Recognition of Indian Sign Language in Live Video, IJACSA, Volume 70– No.19, May 2013.