A Framework on Hand Gesture Recognition Using Fuzzy-Logic with Backpropagation Neural Network Algorithm

[1]Abhishek Patil, [2]Kanchan Pathak, [3]Saurabh Tagalpallewar, [4]Tejas Bharambe
[1,2,3,4]Computer Engineering Department, MIT-COE, Pune, Maharashtra, India

Abstract
This paper presents a method of hand gesture sequence recognition through a multi-feature criteria based on a predefined area in two dimensional space, the presence of skin color, and non-motion for a defined period of time using a web camera. Hand gesture recognition is an active topic of research and development for human computer interfaces. Hand gestures naturally consist of movement and paused motion. This technique focuses on the non-motion features of a hand gesture. It is easy for a human being to recognize the meaning of a common hand gesture but it is difficult for a computer to accomplish this same task. A systematic method was developed to distinguish paused motion from hand movements so that pattern recognition techniques can be effectively utilized to interpret the gesture. We have compared the results of gesture recognition using two algorithms - backpropagation feedforward algorithm and neuro-fuzzy algorithm.

Keywords
Neural Networks, Fuzzy Logic, Gesture Recognition, Fuzzy Clustering, Back-Propagation Feedforward Algorithm

I. Introduction

Gesture is a non-verbal way of effective communication. Gesture recognition is an interface with a system using some particular gestures. Hand gestures are used to communicate with the computer system. Here, the hand gesture is taken as an input and the output associated with that particular hand gesture is generated. Recognizing gestures as input enables computer systems to be more accessible or user friendly for the physically-impaired and makes interaction more exciting in a gaming or 3-D virtual reality environment. The steps involved are:- 1) Image Acquisition 2) Image Pre-processing 3) Feature Extraction 4) Comparison.
To implement hand gesture inputs various methods have been used in the past. Some methods use cameras to capture the shape of the hand and picturise the required inputs to be acted upon in order to get the required interpretation of the gesture [2]. Such vision based systems are the most commonly used systems due to their low cost. The captured images are interpreted using different algorithms like finite state machine, interval mathematics, etc. Some other methods involve the use of sensors [3]. The sensors sense the motion information produced by characters written by the user. The measurement unit records the accelerations and velocities of the motion during hand-writing. These value are then processed using methods like FFT and DCT [4].
In this paper we propose a feature extraction and recognition approach, combining (1) Fuzzy Clustering Method to partition the binarized image in the segmented hand gesture into clusters ; (2) extraction of the cluster distance from the origin and the angle of the cluster with respect to the horizontal axis; (3) Since the neural network needs to be first trained and then tested. For training, pass on these two parameters, generally referred to as training set, initially to the back propagation error based algorithm entirely to train the neural network about a particular gesture.(4) For testing, a input test set which is similar to train set is fed to the neural net workto check the accuracy of the particular gesture by matching the current input set with those in the database.

II. Overall Approach

Gesture is a means of communication through any bodily motion or state which commonly originate from face or hand. Gesture recognition can be done with the artificial neural network. An artificial neuron receives one or more inputs and sums them to produce an required output. The neuron has two modes of operation that is Training Mode and Using Mode . In training mode, the neuron is trained to fire (or not), for particular input patterns. In using mode, when a taught input pattern(paradigm) is detected at the input, its associated output becomes the current output for the next step. (If the input pattern does not belong in the taught list of input patterns, the firing rule is applied in order to determine whether to fire or not).

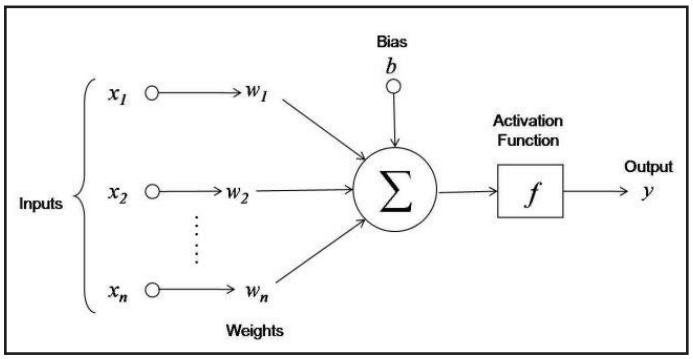

Fig. 1: Structure of Neuron

The neural network is a interconnection of neurons. The neurons are connected to each other by synapses i.e connecting links. Each connecting link has an associated weight. The weight is basically the strength of the connection given by some magnitude. The whole network is divided into layers such as input, hidden and output layer.

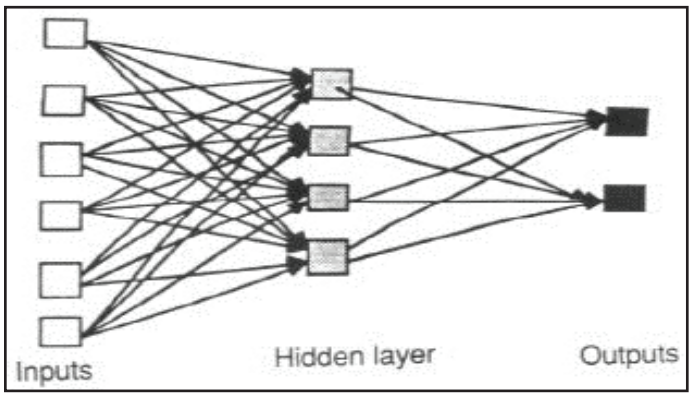

Fig. 2: ANN

The firing rule(for neuron) is an important concept in neural networks for their high flexibility. A firing rule is used to determine how one calculates whether a neuron should fire for any input pattern or not. It relates to all the input patterns, not only the ones on which the node as trained.

A. Image Acquistion

Image acquisition is the capturing of the image using digital cameras. This is the very first step in the process of Gesture Recognition. Cameras used can differ from system to system.

Types of cameras used:

- Webcams can be used to capture images of the hand gesture.
- Similarly there are depth sensing cameras which also can be used. They are used to obtain 3 dimensional images of the hand gesture.
- Stereo cameras can be used which have 2 lenses whose relation to one another is known so as to capture 3D images.
- Wired gloves is another technique which has sensors on it to track the hand gestures.

The main purpose of the pre-processing stage is to:
• Extract just the hand gesture from an image.
• Remove the noises (if any present) and region not required
• Process the extracted image forming a binary image and
• Extract the distinguishable significant features from the processed image, to form a feature set for classification.

B. Image Pre-Processing

1. Segmentation

Segmentation is based on the skin colour. It is used to separate the skin area from the background. The segmentation procedure is not primarily concerned with what the region represents but just with the process of dividing the image. In the simples case (binary images) there are only two regions: a foreground region i.e. the object and a background region. After the segmentation process, the hand region is assigned with white colour and the background is assigned with black colour. Thresholding techniques is done in order to partition the image histogram by using a single threshold, " T ". Segmentation is then done by scanning the image pixel wise and labelling each pixel as object or as background depending on whether its gray level pixel is greater of less than the value of T.

2. Noise Reduction

"Noise" is the commonly-used term to describe visual distortion. Low background light is one major factor that causes noise in images. Noise reduction as the name suggests is the reduction of noise from the images acquired using webcam cameras. The mean filters are just like average filters. They operate on local groups of pixels called neighbourhoods and replace the centre pixel with an average of the pixels in this neighbourhood.

3. Edge Detection

Edge detection operations are based on the idea that edge information in an image is found by looking at the relation a pixel has with its neighbour pixels with widely varying gray levels(binary levels). Ideally an edge is caused by changes in colour or texture or by the specific lighting conditions present during the image acquisition process.

Sobel operator is recognized as one of the best simple edge operators. Horizontal is the row mask, whereas vertical is the column mask. Sobel utilizes two (3x3) masks.

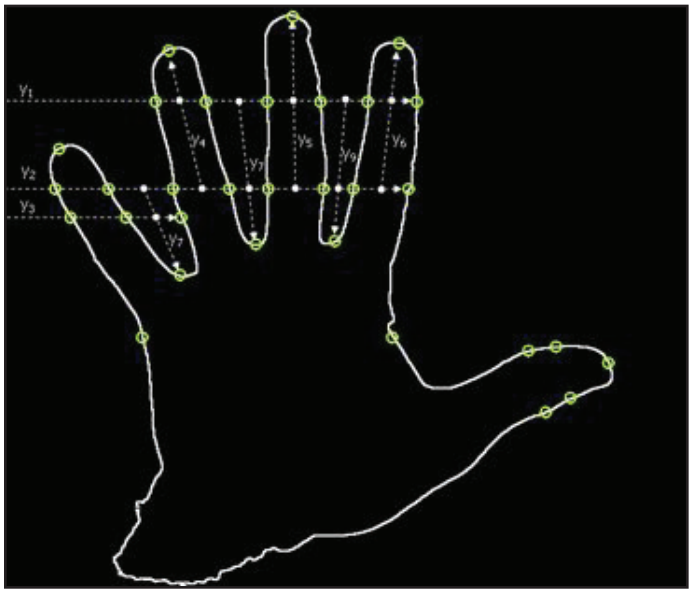

Fig. 3: Edge Detection

These masks are each convolved with the image. At each pixel location there are two numbers S1 corresponding to the result from the row mask and S2 from the column mask these numbers are used to compute two metrics, the edge magnitude and the direction which are define as follows:

Edge Magnitude = $(S1^2 + S2^2)^{1/2}$
Edge Direction = $\tan^{-1}(S1/S2)$

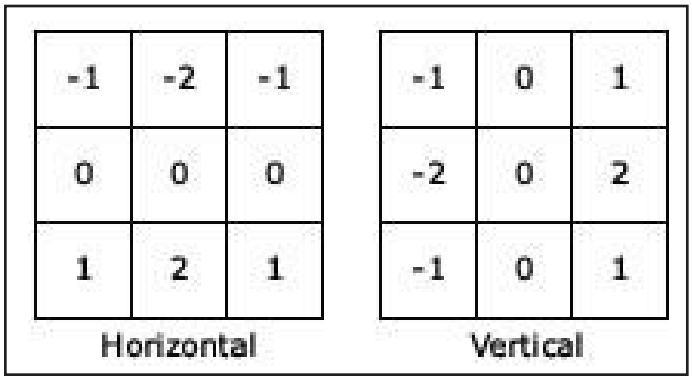

Fig. 4: Sobel Row & Column Mask

C. Feature Extraction

Feature extraction is the process to retrieve the most important data from the raw data. The goal of feature extraction is to discover or find the most discriminate information in the acquired image. For static hand gesture recognition, it is possible to recognition hand posture by identifying and extracting geometric features. Mathematically, a feature is N-dimensional vector with its components by certain analysis of image. The most commonly used visual cues are texture, colour and shape of the hand. The selection of good features is crucial to gesture recognition because hand gestures are rich in shape variation, motion and its textures.

III. Algorithm

Step 1
We resize the image to 300x300 pixels, segment the hand and normalize the coordinates. The main idea is to localize specific gesture regions based on the averaged grayscale values of edge pixels.

Step 2
The given data in the online clustering process, sets consists of input vectors $X = x_1,..., x_p$, which are p points in q-dimensional space. P – no of pixels that is to be processed. Q=3 dimensions of the pixels(x,y) and the value in V grayscale value channel. The algorithm starts with an empty set of clusters. Each new cluster has a cluster radius(Ru) and a cluster centre (Cc). Then start taking the pixel from the origin i.e. the left most top pixel till the end of the screen i.e. right most bottom pixel one by one.

Step 3
The particular may be bounded to a certain cluster or it may lead to building of the new cluster. When a new cluster Ck is created, the current input vector is assigned to be the cluster centre(Cc) and its luster radius(Ru) is initially set to zero. For the assignment of the particular pixel to any cluster it has to be first normalised with the help of $x_{inorm} = x_i - x_{min}/x_{max} - x_{min}$ and the V grayscale value of each pixel need to be considered in the [0-1] range for the membership degree.

Step 4
The cluster is updated or new cluster is created depending upon the threshold value that affects the number of clusters which is given by:

$$D_{ij} = ||x_i - Cc_j|$$

If this is <= atleast one of the threshold of any cluster, it means xi belongs to that cluster. Else it leads to creation of new cluster. The maximum distance from any cluster center to the examples that belong to this cluster is not greater than the threshold value D_{thr}.

Step 5
Dthr affects the number of clusters by updating existing clusters to changing their centre positions, increasing their radii or creating new clusters. Dthr affects the number of clusters by updating the existing clusters changing their centre position. The input vector is bound to each cluster by means of a membership degree (Md), which is number between 0 and 1,and makes clustering more accurate in case of overlapping clusters. The membership degree is assigned as that number of pixels are bound in that cluster with respect to the maximum number of pixels that the cluster can accommodate. We incorporate fuzzy membership degrees and distance to a reference centre for each cluster into clustering process. Once the clustering is done, the output seems like something given in the figure.

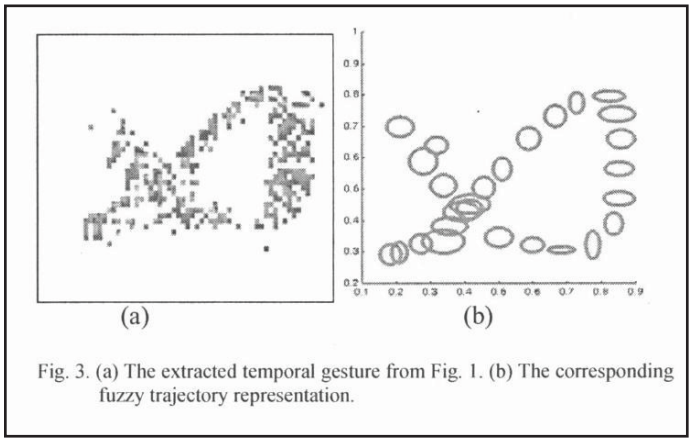

Fig. 3. (a) The extracted temporal gesture from Fig. 1. (b) The corresponding fuzzy trajectory representation.

Fig. 5:

Step 6
The distance from the origin to the cluster centre is noted for every cluster. For every cluster radius and its membership degree are noted.

Step 7
A list of array in the form of [cluster distance from centre : , cluster angle:, cluster membership degree] is maintained. The membership degree is used to filter the number of clusters equal to fixed value 'n' based on the fact that cluster with the maximum membership degree comes first. The clusters are sorted based on their membership degree and first 'n' clusters are filtered. This cluster array is fed to backpropagation algorithm.

Backpropagation Algorithm
Since the output from the above fuzzy process is in terms of vectors. Here, we get three vectors-distances of clusters from the reference, angles of each cluster's centre from the reference and the membership values.

Dataset:

The data from the fuzzy process is gathered and hence,this way a proper dataset is generated.

A dataset (in reference to gesture recognition) is a collection of data of various input patterns and their corresponding obtained outputs.

This dataset has to be split in two parts. One as a training set while other as a test set.

- Training set is used to teach a network to learn and recognize how the output is supposed to be. It can be done by the means of various learning algorithms.
- Test set is to calculate the accuracy of the trained neural network. Depending upon the results of the test data, changes that are needed to be performed on algorithm or network is deduced.

Basically, a dataset is divided into these two sets in some ratio. For eg.70% of a dataset may be used for training the network while the rest 30% for testing the network.

Backpropagation learning algorithm:

One of the most popular Neural Networks algorithms is backpropagation algorithm. It is actually an acronym for "backward propagation of errors". It is a method to train ANN.

Backpropagation is a supervised learning method as it needs to know a desired output for each input value so that it could calculate the error. Moreover, it is a generalization of delta rule for multi-layered feedforward networks and iteratively calculates the gradient for each layer. Backpropagation needs that the transfer (activation) function used by nodes (artificial neurons) be differentiable.

The backpropagation algorithm can be distinguished into two stages : forward propagation and weight updation by backpropagation of error.

1.Forward propagation:

Steps involved in each propagation:
The training set's input pattern (parameters in terms of vectors) is given as input to the input layer of neural network and propagated forward through the hidden layers and lastly the output is computed by the output layer.
In the beginning of training phase of neural network, the weights are randomly assigned to each connection link (synapse).The weights should be a small value.

Since each basic neuron in neural networks has two units:
i) Summation unit
ii) Activation function unit

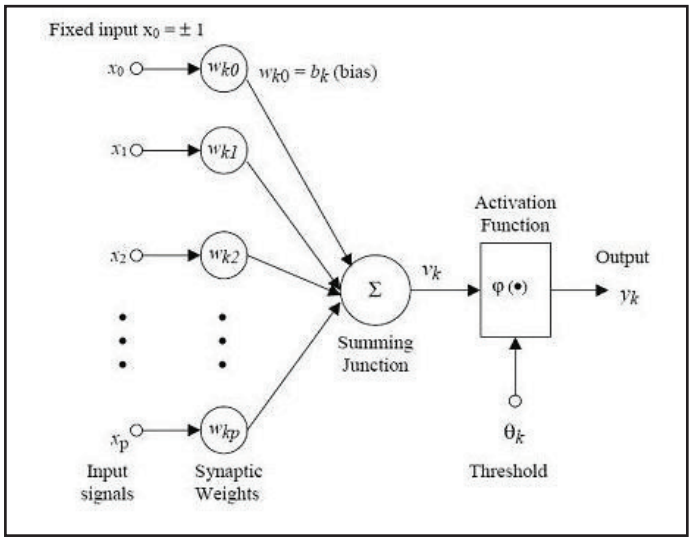

Fig. 6: Units of ANN

In this unit, computation of the weighted sum of all of the inputs is performed and proceeded for further operation in activation unit.
$Y_i = \Sigma x_i w_i + b$
Y_i : output of summation unit
X_i : input given to the neuron
w_i : weight of the connection link for input value
b: bias of the neuron

For each neuron,it has one bias weight with input value 1.Bias weight is always considered when in case there is no input, still the neural network should work properly. It allows the activation function to shift right or left for proper learning.

ii)Activation unit:
The summation received from the summation unit is then provided to the activation function. The result from the activation function produces output for that particular neuron. The activation function can be of any type as per the problem demands. It can be sigmoid function, linear threshold (unit step) function, etc.

$O_i = f(Y_i)$
O_i : output of activation unit of neuron
Y_i : output of summation unit

When the output layer produces an output for the given input pattern, this obtained output is compared with the actual obtained output value. These actual and obtained output values are thus used to calculate the error and used for the next stage i.e. weight updation by propagating this error backwards.

2. Weight updation by back propagating error:

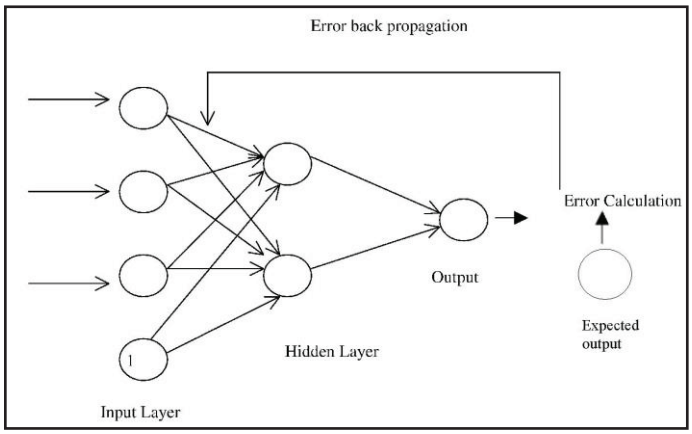

Fig. 6: Back propagation

Since, initially we provide random small weights to the connection links between layers of neuron. This stage is for updating those weights in such a way that they produce accurate and required output in the end.
So, the error calculated is propagated backward to each neuron so that each connection link can update its weight i.e. new weight is determined by adding the previous weight to the error term. The error is the difference between desired output and obtained output

$\delta_i = (t_i - o_i) f(y_i)$
δ_i : error calculated
t_i : desired output value
o_i :obtained output value

So, in backpropagation learning algorithm, the training data trains the neural network to produce accurate output which can be later tested by the test data.

IV. Related Work

Each system for gesture recognition has some shortcomings. In [4] the authors describe in detail the drawbacks of gesture recognition systems. In [5] backpropagation (BP) algorithm is applied for gesture classification, whereas in [6] it was shown that BP performance is inferior to using k-mean based Radial Basis Function neural network. In [6] Otsu gesture segmentation and contour tracking algorithm is employed, which uses Euclidean distance similarity measure.

V. Conclusion

In this paper we propose a lightweight hand gesture recognition system, based on feature extraction and recognition approach, implementing the Evolving Fuzzy Clustering Method and backpropagation algorithm. Hand gestures are a common mode of expressing emotions and information in any conversation. Hand gesture being an important integral part of the conversation even for physically able persons, such hand gesture recognition systems provides marvelous scope in communication for the physically disabled. This paper presents yet another method of recognizing hand gestures.

Since the latest technologies involve high end micro controllers, algorithms can be easily made to recognize the hand gestures based on the inputs to the micro controller and the rules on which the algorithms act upon. The hand features are identified in a new way as grayscale patterns for the edges of neighbor pixels. For gesture recognition we use a formalism of Symbol Relation Grammars to describe a gesture, as well as simple and fast bitwise operations to find the position and orientation of the features. This makes the system open and very easy to be extended with new propositional sentences. The overall approach will be extended to include new types of features and gestures and to investigate the corresponding processing time.

References

[1] Benjam In C. Bedregal, Antonio C. R. Costa And Gracaliz P. Dimuro,"Fuzzy Rule-Based Hand Gesture Recognition", International Federation For Information Processing, Vol. 217, 2006.

[2] Rohit Verma, Ankit Dev,"Vision Based Hand Gesture Recognition Using Finite State Machines And Fuzzy Logic" International Conference On Ultra Modern Telecommunications And Workshop, Oct 2009.

[3] Zhou Ren, Junsong Yuan, Jingjing Meng, Zhengyou Zhang,"Robust Part-Based Hand Gesture Recognition Using Kinect Sensor," Ieee TranSactions On Multimedia, Vol. 15, No. 5, August 2013

[4] M. Hasan, P. Mishra,"Hand Gesture Modeling and Recognition Using Geometry Features: A Review," Canadian Journal On Image Processing And Computer Vision, Vol. 3, No. 1, pp.12-26, 2012.

[5] J. Yoon, J. Min, S. Cho, Enhancing Hand Gesture Recognition Using Fuzzy Clustering-Based Mixture-Of-Experts Model, Proc. Of The 5th Int. Conf. On Ubiquitous Information Management And Communication, Article No. 72, Acm New York, Usa 2011

[6] D. Ghosh, S Ari, "A STatic Hand Gesture Recognition Algorithm Using Kmean Based Radial Basis Function Neural Network," Information, Communications And Signal Processing (Icics) 2011 pp. 1,5, 13-16.