Paper ID: AM-04

Hand Gesture Recognition by Artificial Neural Network

Soumen paul[1], Md. Abu Shahab Mollah[2], Rajibuzzaman[3], Tanvir Ahmad[4], Kazi Toufiqur Rahman[5]
[1,3]Khulna University of Engineering & Technology, Khulna-9203, Bangladesh
[2,4]Ahsanullah University of Science & Technology, Dhaka, Bangladesh
[5]Daffodil International University, Dhaka, Bangladesh
E-mail: sowmen_paul@yahoo.com, shahabeee_06@yahoo.com, tanvir_165@yahoo.com, jeweleee87@yahoo.com

Abstract

This paper represents hand gesture recognition with artificial neural network. Visual Interpretation of gestures can be useful in accomplishing natural Human Computer Interactions (HCI). In this paper we proposed a method for recognizing hand gestures. We have designed a system which can identify specific hand gestures and use them to convey information. In this system we select the feature vectors by Biorthogonal Wavelet Transform. We have used supervised feed-forward neural network based training and Scaled conjugate gradient Backpropagation algorithm for classifying hand gestures into ten categories: A,B,C,D,G,H,I,L,V,Y . This system gives us good performance for recognizing the gestures. We can get up to 82% correct results on a particular gesture set.

Keywords: ANN, HMM, Radon Transformation, Human Computer Interaction, SCG.

1. Introduction

With the development of ubiquitous computing, computer is becoming more and more important in our daily life. Computer applications require more and more unrestricted interaction between human and computers. Hand gesture is frequently used in people's daily life. A gesture is spatio- temporal pattern, which may be static or dynamic or both. One of the most structured sets of gestures belongs to sign language. In sign language, each gesture has an assigned meaning (or meanings). It involves relative flexure of the user's fingers and consists of information that is often too abstract to be interpreted by a machine. Applications of hand gesture recognition widely range from teleoperated control to medicine. For instance, transform of human hand motion for telemanipulation is especially important in hazardous environment. Another important application of hand gesture recognition is to improve the quality of life of the deaf or non-vocal persons through a hand-gesture to speech system. Due to congenital malfunction, disease, head injuries, or virus infections, deaf or non-vocal individuals are unable to communicate with hearing people through speech. Deaf or non-vocal persons use sign language or hand gestures to express themselves. However, most hearing people do not have the special sign language expertise. This is a major barrier between these two groups in daily communication. To overcome this barrier to help those people to integrate into society is a very challenging research area.
The first gestures that were applied to computer interactions date back to the PhD work of Ivan Sutherland [1], who demonstrated Sketchpad, an early form of stroke-based gestures using a light pen to manipulate graphical objects on a tablet display. This form of gesturing has since received widespread acceptance in the human-computer interaction (HCI) community. In the last decade, several methods of potential applications [2], [3], [4] in the advanced gesture interfaces have been suggested but these differ from one to another in their models. Some of these models are Neural Network [2], Fuzzy Systems [3] and HMM [4], [5], [6], [7].Viola uses integral images as Haar wavelet features in rapid object detection [8]. Integral images allow for the fast implementation of box type convolution filters, which makes very fast feature extraction. Farid Parvini and Dennis McLeod presented hand gesture recognition by feature subset selection utilizing Bio-Mechanical characteristics [9]. Hasanuzzaman et al. [10] presented a real-time hand gesture recognition system using skin color segmentation and multiple-feature based template matching techniques. In their method, the three largest skin-like regions are segmented from the input images by skin color segmentation technique from YIQ color space and they are compared for feature based template matching using a combination of two features correlation coefficient and

minimum (Manhattan distance) distance qualifier. Ho-Sub et al. [11] introduced a hand gesture recognition method, which used the combined features of location, angle and velocity to determine the discrete vector that is used as input to HMMs. This method runs over the alphabets (A-Z), numbers (0-9) and six edit commands. Wu [12] developed a hand gesture recognition system for media player control. The system firstly separated the left arm by background subtraction and detected the straight line by both Hough transform and Radon transform.

In this paper we proposed a system where we select the feature vectors by Biorthogonal Wavelet Transformation for recognizing static hand gestures. Here a supervised feed forward neural network has been used with Scaled conjugate gradient Backpropagation learning with a hidden layer having the output nodes representing classes of static hand gesture classification.

2. System Overview

We propose an automatic system that recognizes static hand gesture for alphabets (A,B,C,D,G,H,I,L,V,Y) using Biorthogonal Wavelet Transform. In particular, the proposed system consists of several steps.In the first step images are at first read and then pre-processed. Then noise is removed from the image by using filters. In the next step the proposed method detect the edge from the images and then compute *projections* of an image along specified directions by RT (Radon Transformation). After that, the Biorthogonal Wavelet Transformation is performed on the projections of an image which we get from the RT (Radon Transformation). Then the gestures are trained by supervised feed-forward neural network and then testing is performed. After the completion of testing period the gestures are recognized.

3. Feature Extraction Algorithms

In our proposed system 3 feature extraction algorithms are used. All the algorithms are necessary for recognizing the gestures appropriately.

I. Canny Edge detection Algorithm:

The Canny algorithm uses an optimal edge detector based on a set of criteria which include finding the most edges by minimizing the error rate, marking edges as closely as possible to the actual edges to maximize localization, and marking edges only once when a single edge exists for minimal response. According to Canny, the optimal filter that meets all three criteria above can be efficiently approximated using the first derivative of a Gaussian function.

$$G(x,y) = \frac{1}{2\pi\sigma^2} e^{-\frac{x^2+y^2}{2\sigma^2}} \quad \text{(3.1)}$$

$$\frac{\partial G(x,y)}{\partial x} \;\; \alpha \;\; x e^{-\frac{x^2+y^2}{2\sigma^2}} \quad \text{(3.2)}$$

$$\frac{\partial G(x,y)}{\partial y} \;\; \alpha \;\; y e^{-\frac{x^2+y^2}{2\sigma^2}} \quad \text{(3.3)}$$

The first stage involves smoothing the image by convolving with a Gaussian filter. This is followed by finding the gradient of the image by feeding the smoothed image through a convolution. The 2-D convolution operation is described in the following equation.

$$I'(x,y) = g(k,l) \times I(x,y)$$
$$= \sum_{K=-N}^{N} \sum_{l=-N}^{N} g(k,l) I\,(x-k, y-l) \quad \text{(3.4)}$$

Where: *g (k, l)* = convolution kernel
I(x, y) = original image
I'(x, y) = filtered image
2N + 1 = size of convolution kernel

II. Radon Transformation:

In mathematics, the Radon transform in two dimensions, named after the Austrian mathematician Johann Radon, is the integral transform consisting of the integral of a function over straight lines. The transform was introduced by Johann Radon (1917). Radon further included formulas for the transform in three-dimensions, in which the integral is taken over planes.

Applying the Radon transform on an image $f(x,y)$ for a given set of angles can be thought of as computing the projection of the image along the given angles. The resulting projection is the sum of the intensities of the pixels in each direction, i.e. a line integral. The result is a new image $R(\rho,\theta)$. This can be written mathematically by defining

$$\rho = x\cos\theta + y\sin\theta \quad (3.5)$$

after which the Radon Transformation can be written as

$$R(\rho,\theta) = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} f(x,y)\delta(\rho - xcos\theta - ysin\theta)dxdy \quad (3.6)$$

Here, δ is dirac delta function.

A projection of a two-dimensional function $f(x,y)$ is a set of line integrals. The *radon* function computes the line integrals from multiple sources along parallel paths, or *beams*, in a certain direction. The beams are spaced 1 pixel unit apart. To represent an image, the *radon* function takes multiple, parallel-beam projections of the image from different angles by rotating the source around the centre of the image. The following figure 1 shows a single projection at a specified rotation angle.

Sensors

ρ Rotation Angle Theta

$f(x,y)$

Source

Fig.1. Parallel-Beam Projection at Rotation Angle Theta

For example, the line integral of $f(x,y)$ in the vertical direction is the projection of $f(x,y)$ onto the *x*-axis; the line integral in the horizontal direction is the projection of $f(x,y)$ onto the *y*-axis. The following figure shows horizontal and vertical projections for a simple two-dimensional function.

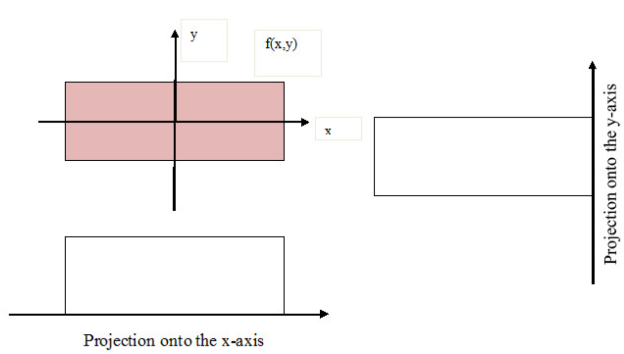

Fig.2. Horizontal and Vertical Projections of a Simple Function

III. Biorthogonal Wavelet:

A biorthogonal wavelet is a wavelet where the associated wavelet transform is invertible but not necessarily orthogonal. Designing Biorthogonal wavelets allows more degrees of freedom than orthogonal wavelets. One additional degree of freedom is the possibility to construct symmetric wavelet functions. The first example of biorthogonal wavelet basis was constructed by Tchamitchian (1987). Before discuss about the biorthogonal wavelet, we will review briefly on orthogonal wavelet.

4. Hand Gesture Recognition System

Hand Gesture Recognition is a complex but one of the benchmark problem in pattern recognition research. In our proposed technique two major steps are to simulate- Image processing and neural network construction. As MATLAB contains *image processing toolbox* and *neural network toolbox* as well, so to perform this work MATLAB was chosen.

The total work is performed in several steps which are image pre-processing, feature extraction, training by neural network, testing and finally recognizing the hand gesture. The whole process can be demonstrated by the following block diagram.

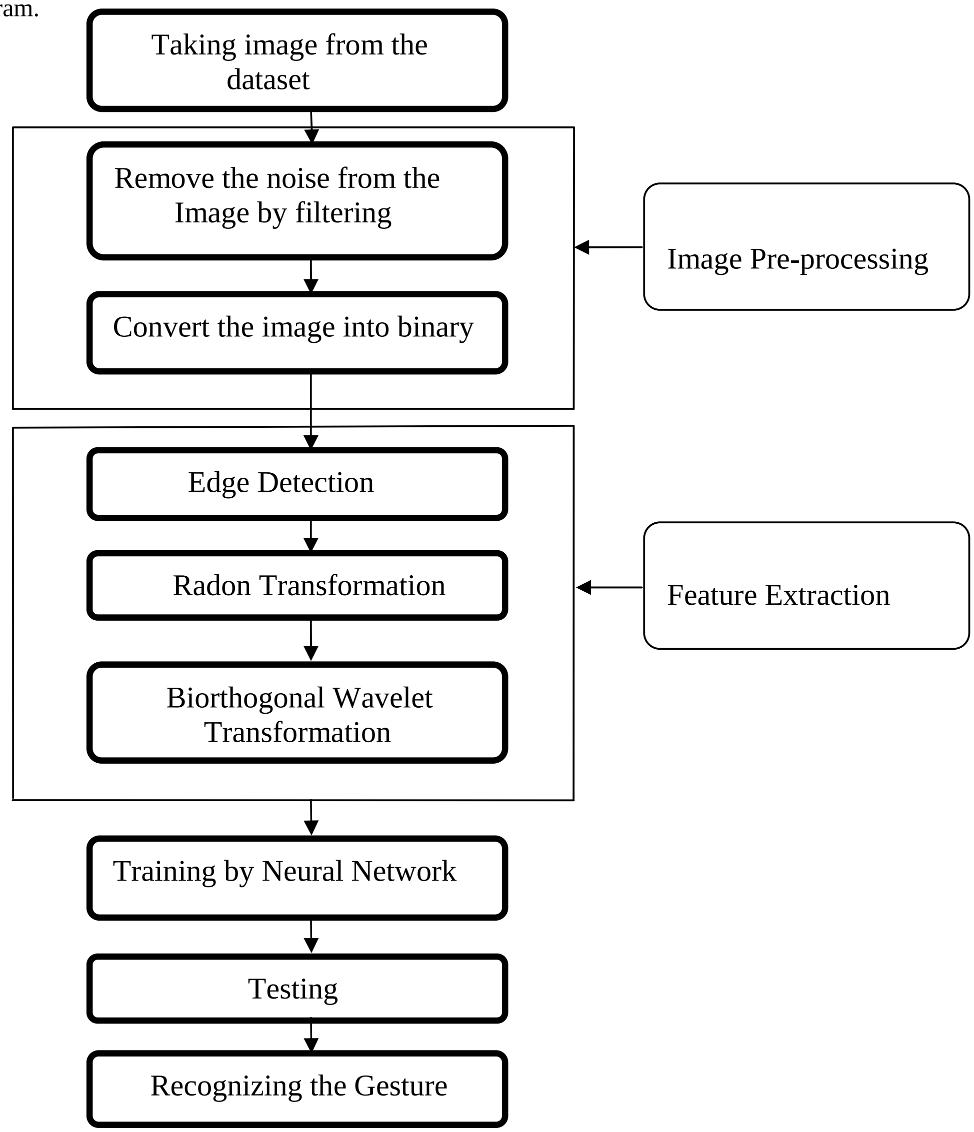

Fig.3. Block diagram of Hand Gesture Recognition System

I. Image Pre-processing:

Before feature extraction of an image, image pre-processing is a necessary step. Image pre-processing is used for operations on images at the lowest level of abstraction. The pre-processing do not increase image information content but decrease it if entropy is an information measure. For example as Histogram equalization, it modifies the brightness and contrast of the image, making the image look clearer.

In our proposed system image pre-processing is combined of several steps. Our image database [14] was already in the gray scale format, so we did not need to convert in gray scale format. Considering the uniform background is white, first of all we resize all the images. Then we remove the noise from them by Gaussian lowpass filter using gaussian function and *imfilter* function. Converting the images into the binary scale by i*m2bw* function we finished our image pre-processing task.

II. Feature Extraction:

Feature Extraction is a very important task for any system based on artificial neural network. Selecting good feature increase performance of the system. In our proposed system feature extraction is done by the following three steps.

A. ***Edge Detection****:* In an image, an edge is a curve that follows a path of rapid change in image intensity. Edges are often associated with the boundaries of objects in a scene. Edge detection is used to identify the edges in an image.

In this system we detect the edges of an image by Canny edge detection algorithm, because this edge detector performs better than reference algorithms.

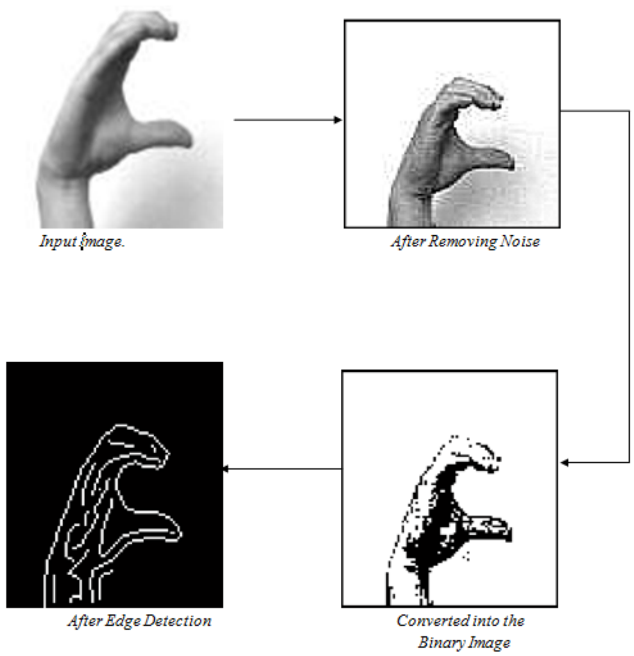

Fig. 4. Edge detection by Canny Edge Detection Algorithm

B. ***Radon Transformation****:* After the detecting the edges of the image, computation of the multiple parallel-beam projections of the image from different angles by rotating the source from 0° to 180° was performed around the center of the image.

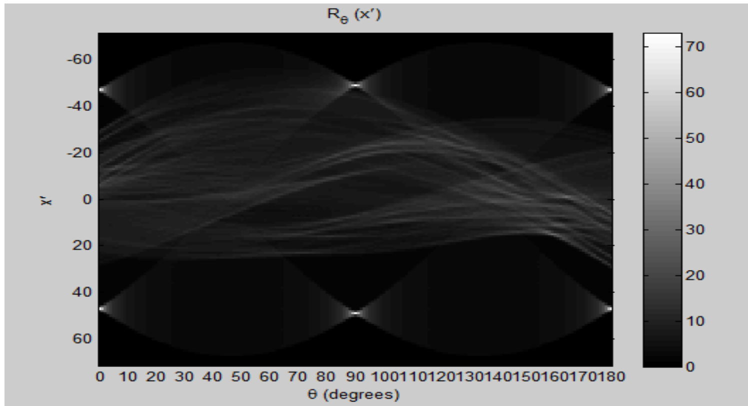

Fig. 5. Plotting of Radon Transformation of an image at angles from 0° to 180°.

In Figure 5 we see that some sine waves appears, it is because the Radon transform of a Dirac delta function(**δ**) is a distribution supported on the graph of a sine wave. Consequently the Radon transform of a number of small objects appears graphically as a number of blurred sine waves with different amplitudes and phases. For that reason the Radon transform data is often called a *sinogram.*

C. ***Biorthogonal Wavelet Transform****:* By performing biorthogonal wavelet transformation on the computed projection which we obtained from the Radon Transformation, the final feature for our recognition system is selected. We used biorthogonal wavelet 3.7 in this system.

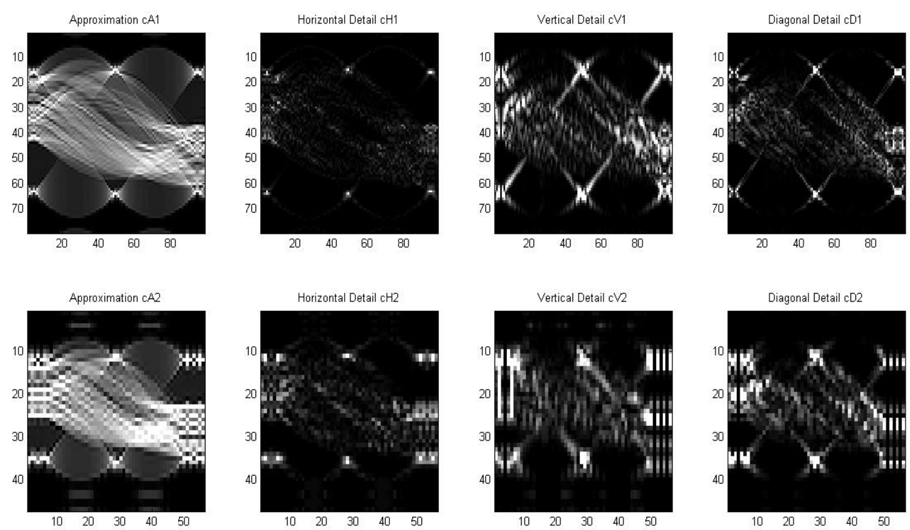

Fig.6. *After Biorthogonal wavelet transform*

III. ***Training By Neural Network:***

For training the feature by neural network, the most important portion is that to create a neural network first and to feed the inputs in that network. In MATLAB for creating a feedforward neural network a built-in function named *newff* was used here. For training the network with scaled conjugate gradient backpropagation trainscg function is used as a network training function.

The performance function is specified according to error calculation. In our program we used *mse (mean squared error)* performance function. The error is calculated as the difference between the target output and the network output. Then the mean squared error is calculated.

$$mse = \frac{1}{Q}\sum_{k=1}^{Q} e(k)^2 = \frac{1}{Q}\sum_{k=1}^{Q}(t(k) - a(k))^2 \quad (4.1)$$

A very important term for the learning procedure is Epoch. The term epoch means a single pass through the entire learning process. When the input vector is pass through the learning procedure the entire process then said to be an epoch. In our program the maximum number of epoch is assigned to 10000. More epochs consume more time for learning. If the number of epoch is less, the learning might be incomplete. Then the probability of error increases.

The following parameters we have used for neural network in our program

Number of epoch: 10,000

Number of hidden neurons: 10

Performance Goal: 0.00001

Learning rate: 0.1

Epochs between showing progress: 50

Training Time: 2s

5. Experimental Result:

The proposed method was tested on five different users showing ten gestures such as A,B,C,D,G,H,I,L,V,Y. Table 1 shows the recognition results of 10 gestures of 5 users.

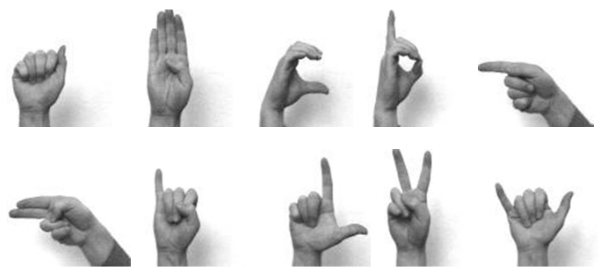

Fig.7. Finger spelled Alphabet (Top row [A, B, C, D, and G], Bottom row [H, I, L, V, Y]).

Table 1: Recognition rate of hand gestures using proposed method for 10 gestures and 5 users

Gesture / User	A	B	C	D	G	H	I	L	V	Y	%
1	9/10 (90%)	8/10 (80%)	7/10 (70%)	7/10 (70%)	7/10 (70%)	6/10 (60%)	8/10 (80%)	8/10 (80%)	6/10 (60%)	6/10 (60%)	72%
2	8/10 (80%)	6/10 (60%)	8/10 (80%)	7/10 (70%)	7/10 (70%)	8/10 (80%)	8/10 (80%)	5/10 (50%)	7/10 (70%)	8/10 (80%)	72%
3	8/10 (80%)	5/10 (50%)	6/10 (70%)	8/10 (80%)	9/10 (90%)	8/10 (80%)	6/10 (60%)	6/10 (60%)	8/10 (80%)	7/10 (70%)	71%
4	9/10 (90%)	10/10 (100%)	10/10 (100%)	7/10 (70%)	7/10 (70%)	9/10 (90%)	5/10 (50%)	6/10 (60%)	7/10 (70%)	6/10 (60%)	76%
5	7/10 (70%)	6/10 (60%)	9/10 (90%)	5/10 (50%)	9/10 (90%)	8/10 (80%)	9/10 (90%)	8/10 (80%)	6/10 (60%)	7/10 (70%)	74%
Total	82%	70%	80%	68%	78%	78%	72%	66%	68%	68%	

We have plotted the Iteration versus Error Curve for the different users of 10 gestures.

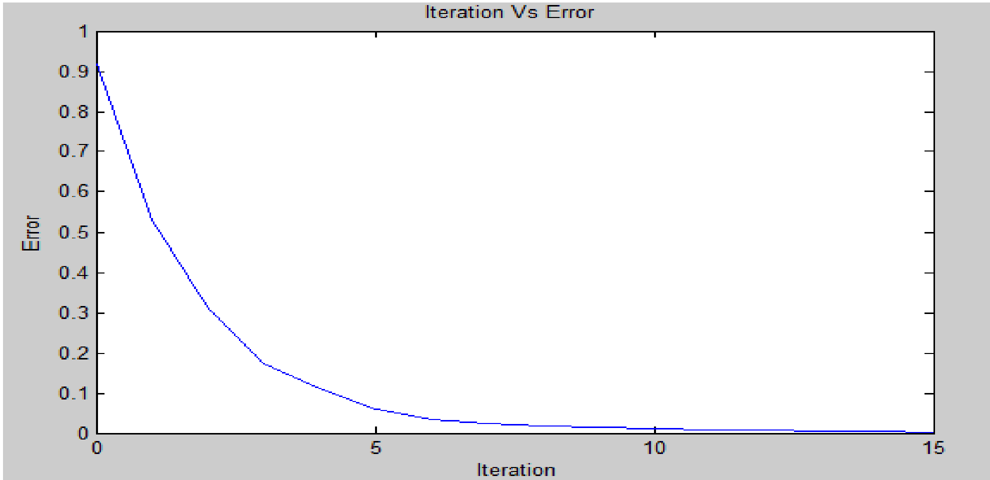

Fig.8. Iteration versus error curve for user 1

The Fig.8 shows that with iteration the error reduces. For other users the error also reduces as the iteration increases. For convenience here another curve is shown for user 3.

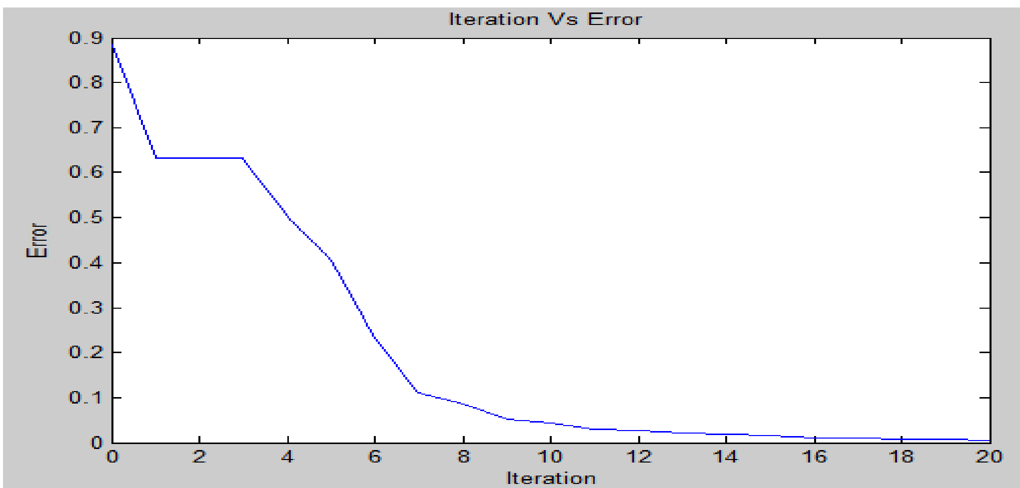

Fig.9. Iteration versus error curve for user 3

6. Discussion:

Artificial Neural Networks (ANN) has been applied to an increasing number of real-world problems of considerable complexity. Their most important advantage is in solving problems that are too complex for conventional technologies, i.e., problems that do not have an algorithmic solution or for which an algorithmic solution is too complex to be found. Here we applied ANNs to the problem of differentiating between the 10 static hand gestures with very good results. In order achieve high accurate result; we applied radon transformation and biorthogonal wavelet transformation formation for selecting the feature. For faster training we used scaled conjugate gradient backpropagation .The accuracy can be achieved up to 82% for a particular gesture set.

7. Future Work:

We only recognized and classify the static hand gestures, so dynamic hand gesture recognition will be more useful for any of this related applications. The future work also will focus on the advanced feature-level algorithms for improving more accuracy of hand gesture recognition

8. References:

[1] Sutherland, I. E. Sketchpad: "*A man-machine graphical communication system".* In: Proceedings of the AFIPS Spring Joint Computer Conference 23. pp. 329–346,1963.

[2] X. Deyou, "*A Network Approach for Hand Gesture Recognition in Virtual Reality Driving Training System of SPG"*, International Conference ICPR, pp. 519-522, 2006.

[3] E. Holden, R. Owens and G. Roy, *"Hand Movement Classification Using Adaptive Fuzzy Expert System*", Journal of Expert Systems, Vol. 9(4), pp. 465-480, 1996.

[4] M. Elmezain, A. Al-Hamadi and B. Michaelis, *"Real-Time Capable System for Hand Gesture Recognition Using Hidden Markov Models in Stereo Color Image Sequences"*, Journal of WSCG, Vol. 16, pp. 65-72, 2008.

[5] M. Elmezain, A. Al-Hamadi and B. Michaelis, "*A Novel System for Automatic Hand Gesture Spotting and Recognition in Stereo Color Image Sequences"*. Journal of WSCG, Vol.17,No. 1, pp. 89-96, 2009.

[6] M. Elmezain, A. Al-Hamadi, J. Appenrodt and B. Michaelis, "*A Hidden Markov Model-Based Continuous Gesture Recognition System for Hand Motion Trajectory"*. International Conference on Pattern Recognition (ICPR), pp. 519-522, 2008.

[7] M. Elmezain, A. Al-Hamadi and B. Michaelis, *"Spatio- Temporal Feature Extraction-Based Hand Gesture Recognition for Isolated American Sign Language and Arabic Numbers"*. IEEE Symposium on ISPA, pp. 254-259, 2009.

[8] P. Viola and M Jones, *"Rapid object detection using a boosted cascade of simple features,"* in Proceedings of Computer Vision and Pattern Recognition. Hawaii,U.S., 2001, pp. 511–518.

[9] Farid Parvini and Dennis McLeod, *"Feature Subset Selection Utilizing BioMechanical Characteristics for Hand Gesture Recognition"* 978-1-4244-4134-1/09/ 2009 IEEE.

[10] Md. Hasanuzzaman, V. Ampornaramveth, Tao Zhang, M.A. Bhuiyan ,Y. Shirai and H. Ueno, *"Real-time Vision-based Gesture Recognition for Human Robot Interaction".* In the Proceedings of the IEEE International Conference on Robotics and Biomimetics, Shenyang China ,2004.

[11] Y. Ho-Sub, S. Jung, J. B. Young and S. Y. Hyun, *"Hand Gesture Recognition using Combined Features of Location,Angle and Velocity",* Journal of Pattern Recognition, Vol. 34(7), pp. 1491-1501, 2001.

[12] Y. M. Wu, *"The implementation of gesture recognition for media player system,"* Master Thesis of the Department of Electrical Engineering, National Taiwan University of Science and Technology, Taipei, Taiwan, 2009.

[13] Martin Fodslette Moller. *"A Scaled Conjugate Gradient Algorithm For Fast Supervised Learning. Neural Networks, 6:525-533, 1993".*

[14] Sebastien Marcel – *"Gesture Database Web Page.htm"*

[15] *"Digital image processing using Matlab"* by Rafael C. Gonzalez, Richard E. Woods, Steven L. Eddins.

[16] *"Neural Networks"* by Simon Haykin. *"http://ece2day.blogspot.com/2012/10/neural-networks-comprehensive.html".*

[17] *"http://cas.ensmp.fr/~chaplais/wavetour_presentation/Wavetour_presentation_US.html."*

[18] *"Hand Pose Estimation for American Sign Language Recognition*" by Jason Isaacs and Simon Foo ,2004 IEEE

[19] Duong Van Hieu; Supot Nitsuwat *"Image Preprocessing and Trajectory Feature Extraction based on Hidden Markov Models for Sign Language Recognition",* Ninth ACIS International Conference on Software Engineering, Artificial Intelligence, Networking, and Parallel/Distributed Computing,2008,Phuket, Thailand.

[20] Mahmoud Elmezain, Ayoub Al-Hamadi, Bernd Michaelis *"Improving Hand Gesture Recognition Using 3D Combined Features"*, 2009 Second International Conference on Machine Vision, 28-30 Dec.2009,Dubai,pp. 128-132.