International Journal on Recent and Innovation Trends in Computing and Communication
Volume: 2 Issue: 5
ISSN: 2321-8169
1305 – 1311

Human-Computer Interaction using Hand Gestures

Prof. Sachin A Urabinahatti.
Dept of Computer Science.
TKIET, WARANANAGAR.
Kolhapur, India.
saurabinahatti@tkietwarana.org

Prof. Apeksha S Bandekar.
Dept of Computer Science.
STE,s Smt.Kashibai Navale Coll Of Engg
Pune, India.
apekshagadkar@gmail.com

***Abstract*--**With the ever-increasing diffusion of computers into the society, the present popular mode of interactions with computers (mouse and keyboard) will become a bottleneck in the effective utilization of information flow between the computers and the human. The use of hand gestures provides an attractive alternative to cumbersome interface for human-computer interaction (HCI). The hand can be used to communicate, with much more information by itself compared to computer mouse, joysticks, etc. allowing a greater number of possibilities for computer interaction. Developing new techniques for human-computer interaction is very challenging, in order to use hands for interaction, it is necessary to be able to recognize them in images. In this project, a robust hand gesture recognition system is designed and presented for recognizing static gestures based on Zernike moments (ZMs) for feature extraction & Artificial Neural Network (ANN) for training & classification of gestures. The proposed system is able to recognize the gesture irrespective of the angles in which the hand gesture image is captured, which makes the system more flexible with uniform background.

Keywords*: Hand Gesture Recognition, Human Computer Interaction, Artificial Neural Network, Zernike Moment*

*****

1. INTRODUCTION

The way humans interact with computers is constantly evolving, with the general purpose i.e., being to increase the efficiency and effectiveness by which interactive tasks are completed. The purpose of this review is to introduce the field of gesture recognition as a mechanism for interaction with computers. Gestures are expressive, meaningful body motions involving physical movements of the finger, hands, arms, head, face, or body with the intent of: 1) conveying meaningful information or 2) interacting with the environment [13]. A hand gesture is defined as a dynamic gesture if movement refers to a sequence of hand postures connected by contiguous motions over a short time span, such as waving good-bye; hand gesture is defined as static gesture if there is no involvement of movements in the hand posture. Hand gesture recognition finds applications in varied domains including virtual environments, smart surveillance, sign language translation, medical systems etc. Hand gestures are an attractive method for communication with the deaf and dumb. Another important application area is that of vehicle interfaces. Hand gesture is used to control an event like navigation of slides in Power Point Presentation i.e., during a presentation, the presenter does not have to move back and forth between computer and screen to select the next slide. Hand Gestures can be used for remote controls for television sets, stereos and room lights. Household robots could be controlled with hand gestures.

2. PROPOSED APPROACH

With the development of ubiquitous computing, current user interaction approaches with keyboard, mouse and pen are not sufficient. Due to the limitation of these devices the usable command set is also limited. Direct use of hands can be used as an input device for providing natural interaction. This work aims to develop a real-time system capable of understanding commands by hand gestures, and expand the ways that people are able to interact with their computers. The work here is summarized as detecting and recognizing hand gesture from real time gesture images and with the help of neural network analyzes the efficiency of the network for the hand gesture recognition problem. Further a sample application of controlling slides through hand gesture is modeled here.

This approach is based on three main steps: Capture hand image, Preprocessing & segmentation, Feature extraction and Gesture recognition.

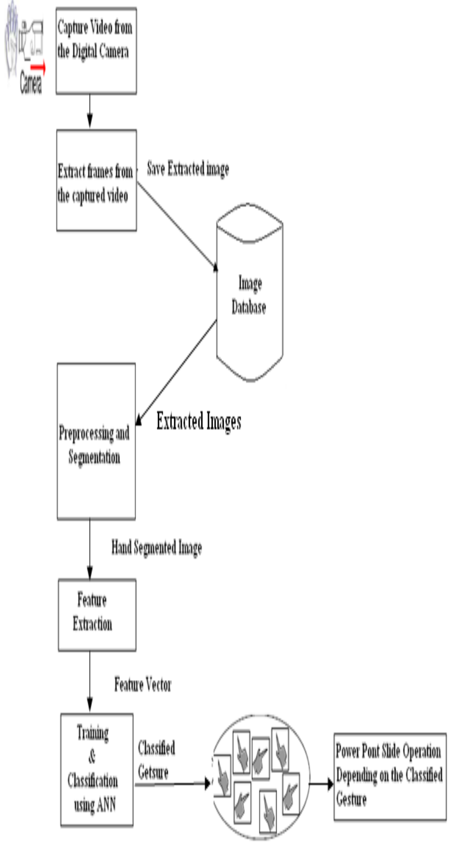

Figure 1: Block Diagram of Proposed System

3. IMPLEMENTATION

Algorithm for the proposed approach
Input: A static gesture image.
Output: The recognized static gesture image makes the system operate.
Step 1: Capture video of hand from the digital camera by converting the video from RGB to gray scale & perform histogram equalization to increase the contrast of video. Extract video frames and store in the image database.
Step 2: Read the frames from stored database; preprocess the image for smoothening and noise removal using median filtering. Segment the input gesture image into the binary hand silhouette using edge detection.
Step 3: Compute the Zernike moments **(**ZMs) in order to extract features considering 11 orders of moments, Store the extracted features in feature vector for training.
Step 4: Build the Feed forward Backpropagation neural network for training & classification of gestures.
Step 5: Once the gesture is classified, the PowerPoint slides is operated according to the meaning specified to the classified gesture.
Step 6: End.

3. 1. Image Database

The creation of the image database with all the images extracted from the video that would be used for training and testing. The construction of an image database is clearly dependent on the application. The image database in the proposed approach consists of 118 image samples with uniform background with a distance or 1 or 2 feet from the camera in light environment; where 59 image sample for each gesture 1 and gesture 2. The image database itself is responsible for the better efficiency of the NN as it is that decides the robustness of the algorithm. In the proposed approach the gesture images The FOS transfer model can be characterized by direct communication between the technologies i.e. gesture 1 and gesture 2 are used for navigation of PowerPoint slides where gesture 1 indicates to move the slide next and gesture 2 indicates to move the slide behind. The figure 2 shows some of the images of gesture 1 and gesture 2 images stored in the data base for training and testing of the NN.

Figure 2: The Image Database

The example images below in the figure 3 are the training testing images.

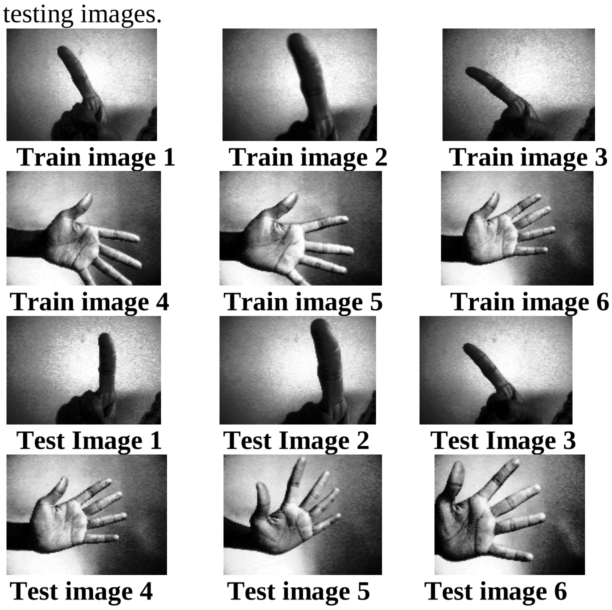

Figure 3: Train – Test Images

3.2. Feature Extraction

The aim of this phase is to find and extract features that can be used to determine the meaning of a given gesture. The PowerPoint application in the proposed approach is based on shift and rotation invariant input gesture image because shift and rotation invariant features lead to a better recognition of hand gestures even if the hand gesture is captured in a different angle. Hence Zernike moments are used to calculate feature set due to its shift and rotation invariant feature. The proposed approach considers 11 orders of moments to extract the feature from the segmented image.

3.2.1. Zernike moments (ZM)

3-d or 2-d objects are generally recognized with the help of their shapes and most of the real time objects have irregular shapes. Hence they cannot be properly described with the help of regular shape descriptors like circularity, linearity and so on. Hence the proposed approach adopts Zernike moments. The moments are higher space feature vector and are generally of order N. The more order of moments are considered, the better the recognition probability. If any image is assumed to be an object, its descriptors are known as feature vectors.

The Zernike polynomials are a set of complex, orthogonal polynomials defined over the interior of a unit circle x2 + y2 = 1.Zernike moment of order n and repetition m is defined as:

$$Z_{nm} = n{+}1/\pi \int\int_{x2+y2\leq 1} V\,nm(\rho,\Theta) f(x,y)\,dxdy \quad (1)$$

Where:
f(x,y) is the image intensity at (x,y) in Cartesian coordinates, n is a non-negative integer, m is an integer such that $n–|m|$ is even positive integer and $|m|\leq n$, □ □ □ x2 □ □y2
Θ is the angle between vector ρ and the x -axis in a counter clockwise direction.

 ISSN: 2321-8169
1305 – 1311

The form of orthogonal Zernike basis polynomials, *Vnm*(ρ, θ)
Vnm(ρ, θ) is a complex conjugate of *Vnm*(ρ, θ) = $R_{nm}(\rho)e^{-jm\Theta}$ in polar coordinates

(ρ,Θ) and j=$\sqrt{-1}$

The polar coordinates polar coordinates (ρ,Θ) in the image domain are related to Cartesian coordinates (x,y) as x= $\rho \cos(\Theta)$ and y= $\rho \sin(\Theta)$.
$R_{nm}(\rho)$ is a radial defined as follows:

$$R_{nm}(\rho) = \sum_{s=0}^{n-m/2} \frac{(-1)^s\ [(n-s)!\rho^{n-2s}]}{S!\left(n+\frac{|m|}{2}-s\right)\left(n-\frac{|m|}{2}-s\right)} \qquad (2)$$

To calculate the Zernike moments of an image f(x,y), the image is first mapped onto the unit disk using polar coordinates, where the center of the image is the origin of the unit disk. Pixels falling outside the unit disk are not used in the calculation. Because Z_{mn} is complex, we use the Zernike moments modules Z_{mn} as the features of shape in the recognition of patterns. The magnitude of Zernike moments has rotational invariance property. An image can be better described by a small set of its Zernike moments than any other type of moments such as geometric moments, Legendre moments, and complex moments in terms of mean-square error. Zernike moments do not have the properties of translation invariance and scaling invariance. The way to achieve such invariance is image translation and image normalization before calculation of Zernike moments [29].

3.3. Gesture Recognition & Classification

The recognition process consists of two phases, training and classification. To recognize the gesture and classify recognized gesture Artificial Neural Network is used, where the datasets are trained by using feed forward back propagation algorithm.

3.3.1 Feed forward Backpropagation Neural Network

Feed Forward neural network is an artificial neural network which is used to train the network, where connections between the units do *not* form a directed cycle. In this network, the information moves in only one direction, forward, from the input nodes, through the hidden nodes (if any) and to the output nodes. There are no cycles or loops in the network.

The term "Backpropagation" describes how this type of neural network is trained. Backpropagation is a form of supervised training. In back-propagation, there are two phases in its learning cycle, one to propagate the input pattern through the network and the other to adapt the output, by changing the weights in the network. It is the error signals that are back propagated in the network operation to the hidden layer(s). The back propagation uses following values:

Transfer function: log-sigmoid in hidden layer and linear transfer in output layer. Epochs: 4500, describes the number of iterations.

The figure 4 shows the feed forward back propagation neural network in proposed approach which consist of 11 input nodes which takes the input from the feature vectors calculated using Zernike moments by taking 11 order of moments, Since i=11 input neurons the hidden neurons will be h=11*11 hidden neurons= 121 hidden neurons, o=2 output neurons.

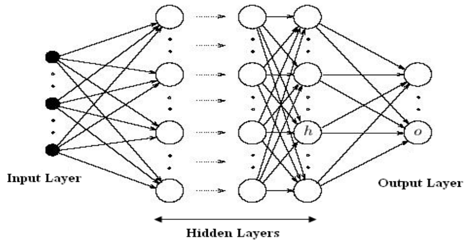

Figure 4: FFBP Neural Network training phase

4. EXPIREMENTAL ANALYSIS & RESULTS

4.1. Experimental Analysis

Experimental Analysis WRT Number of features v/s Neural Network efficiency

Number of Features	Recognition Rate for Gesture 1 (%)	Recognition Rate for Gesture 2 (%)	Overall Neural Network Efficiency (%)
2	66	52	59
3	82	70	76
4	77	51	64
5	56	48	52
6	70	55	62.5
7	85	70	77.5
8	79	59	69
9	71	73	72
10	60	55	57.5
11	94	71	82.5

Table 1: Number of features v/s Neural Network efficiency.

The table 1 shows the dependency of the Neural Network efficiency on the number of features. Number of features represents the n order of Zernike moments, where n is number of feature vector for per training set. The efficiency of the network is optimum when there are 11 order of moments i.e., 11 feature vector. The below figure 5 shows the graphical representation of analysis with respect to Number of Features v/s Neural Network Efficiency which shows the neural network is optimum when there are 11 order of Zernike moments.

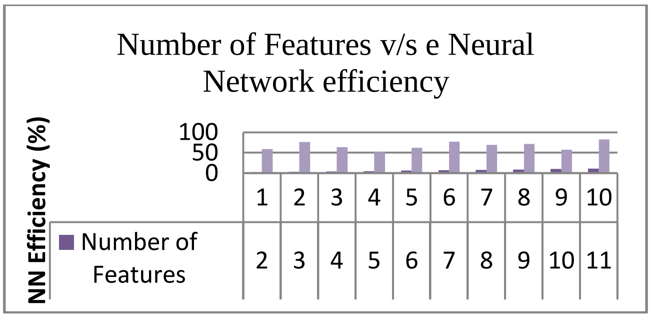

Figure 5: Graphical Analysis for Number of Features v/s NN efficiency

Experimental Analysis WRT Number of Hidden Neurons v/s Neural Network efficiency.

Number of Hidden Neurons	Recognition Rate for Gesture 1 (%)	Recognition Rate for Gesture 2 (%)	Overall Neural Network Efficiency (%)
80	84	61	72.5
100	84	64	74
121	94	71	82.5

Table 2: Number of Hidden Neurons v/s Neural Network efficiency.

The table 2 shows the dependency of the efficiency on the number of hidden layers. Number of hidden layer represents number of states of the neurons in the network. The efficiency of the network is optimum when there are at least n x n numbers of hidden layers. The n here represents number of features per training set. The below figure 6 shows the graphical representation of analysis with respect to Number of Hidden Neurons v/s Neural Network Efficiency which shows the network is optimum when 121 hidden neurons are considered.

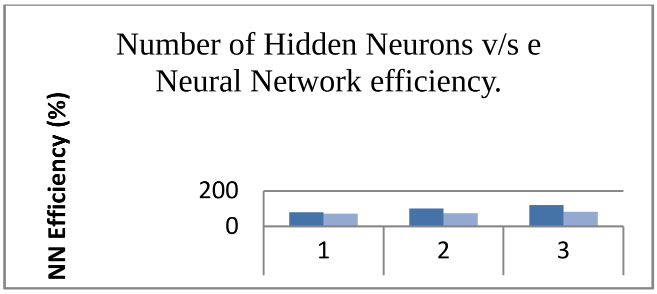

Figure 6: Graphical Analysis for Number of Hidden Neurons v/s NN efficiency

Experimental Analysis WRT Termination Error Rate v/s Neural Network efficiency

Termination Error rate (ms)	Recognition Rate for Gesture 1 (%)	Recognition Rate for Gesture 2 (%)	Overall Neural Network Efficiency (%)
0.1	75	56	65.5
0.01	84	64	74
0.001	77	52	64.5
0.0001	67	65	66
0.00001	94	71	82.5

Table 3: Termination Error Rate v/s Neural Network efficiency

The table 3 shows the dependency of the efficiency on the termination error rate. Termination error rate represents the maximum tolerable error in classifying the values in a neural network. The efficiency of the network is optimum for more termination rate, better is the performance of the neural network. The below figure 7 shows the graphical representation of analysis with respect to Termination Error Rate v/s Neural Network Efficiency which shows the network is optimum when termination error is set to 0.00001.

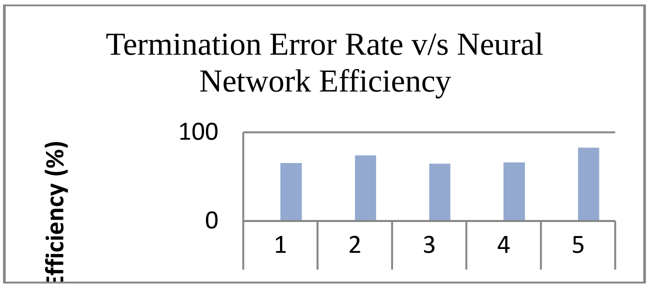

Figure 7: Graphical Analysis for Termination error rate v/s NN efficiency

Experimental analysis with respect to gestures with cluttered background

The network is trained on 59 samples of gestures from which 30 samples are used for training and 29 for testing. Samples with cluttered background are taken into consideration. The Below table 4 shows the recognition rate for gesture 1 and gesture 2 by setting the parameters as specified in table 5.8. Figure 8 shows the Performance of Neural Network with gestures of cluttered background.

Gesture	Recognized Samples	Misclassified samples	Recognition Rate (%)
1	18	1	61
2	13	8	44

Table 4: Recognition rate of the gestures with cluttered background

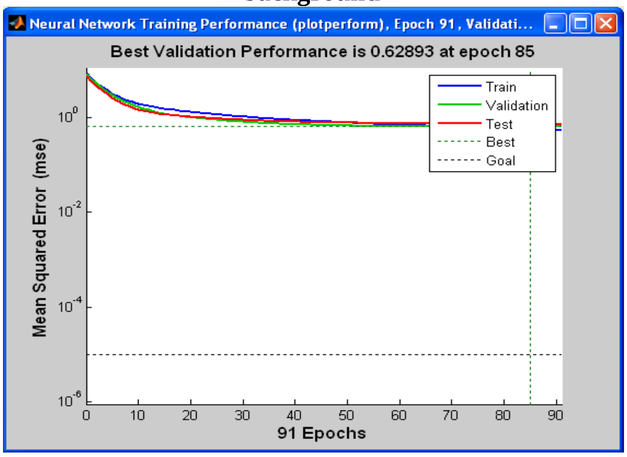

Figure 8: Performance of Neural Network with gestures of cluttered background

Experimental analysis with respect to gesture input in dark environment at distance of 1 meter from the camera

The network is trained on 59 samples of gestures from which 30 samples are used for training and 29 for testing. The Below table 5 shows the recognition rate for gesture 1 and gesture 2 by setting the parameters as specified in table 8. Figure 9 shows the Performance of Neural Network with gesture input dark environment at distance of 1 meter from the camera.

Gesture	Recognized Samples	Misclassified samples	Recognition Rate (%)
1	18	1	61
2	16	8	54

Table 5: Recognition rate of the gesture with gesture input in dark environment at distance of 1 meter from the camera.

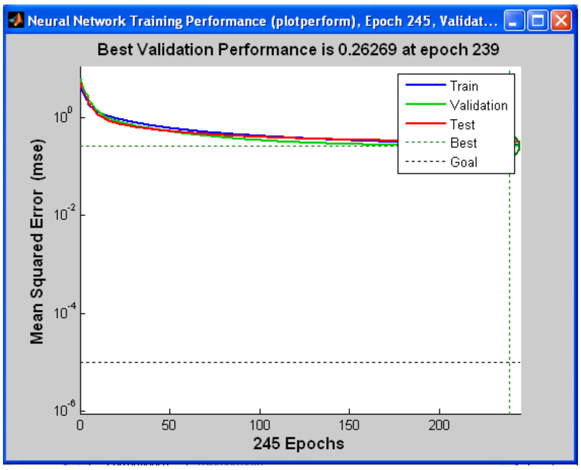

Figure 9: Performance of Neural Network with gesture input dark environment at distance of 1 meter from the camera.

Experimental analysis with respect to gesture input in light environment and with a distance of 1 meter from the camera

The network is trained on 59 samples of gestures from which 30 samples are used for training and 29 for testing. The Below table 6 shows the recognition rate for gesture 1 and gesture 2 by setting the parameters as specified in table 8. Figure 10 shows the Performance of Neural Network with gesture input in light environment at distance of 1 meter from the camera.

Gesture	Recognized Samples	Misclassified samples	Recognition Rate (%)
1	19	1	64
2	16	8	54

Table 6: Recognition rate of the gesture with gesture input in light environment at distance of 1 meter from the camera.

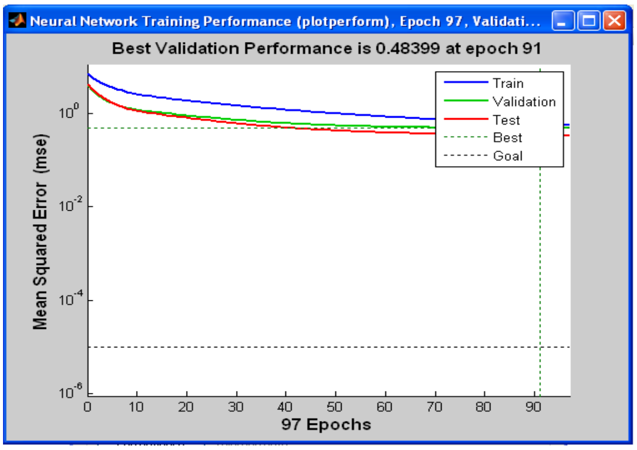

Figure 10: Performance of Neural Network with gesture input in light environment at distance of 1 meter from the camera.

4.2. Result

Finally, the network is trained on 59 samples of gestures from which 30 samples are used for training and 29 for testing. Samples with uniform background in light environment with minimum distance o 1 or 2 feet are taken into consideration. The Below table 7 shows the recognition rate for gesture 1 and gesture 2 by setting the parameters as specified in table 8.

Gesture	Recognized Samples	Misclassified samples	Recognition Rate (%)
1	28	1	94
2	21	8	71

Table 7: Recognition rate of the gestures with uniform background

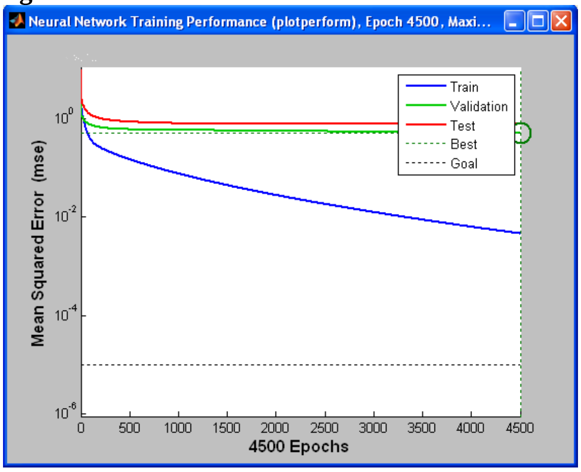

Figure 11: Performance of Neural Network with gestures of uniform background

Structure of Neural Network for proposed approach

Number of Input Neurons	11
Number of output Neurons	2
Hidden Layers	11
Hidden Layer Neurons (11*11)	121
Iterations	4500
Termination Error Rate (ms)	.00001
Gradient	0.02
Recognition Rate of Neural Network (%)	82.5

Table 8: Neural Network Design

From the above conducted analysis the table 8 clearly shows that the performance of neural network is not only depending upon the number features; number of hidden neurons and the termination error rate but also depends on the quality of gesture image. Hence an optimization must be tested with number of feature values, number of hidden neurons and the termination error rate in various different input conditions in order to correctly classify samples to their corresponding classes.

The system performance can be evaluated based on its ability to correctly classify samples to their corresponding classes. Hence the above shown experimental analysis shows that the efficiency of the network is better when number of features are 11, number of hidden neurons are 121, termination error rate is .00001 and images with uniform background in light environment with minimum distance of 1 or 2 feet between the input gesture object and the camera, the performance of the NN is depicted in figure 17 which shows the training has reached minimum gradient value of 0.02 with 4500 iterations.

ISSN: 2321-8169
1305 – 1311

Recognition Rate

The recognition rate can be defined as the ratio of the number of correctly classified samples to the total number of samples and can be given as

Recognition rate = number correctly classified signs/Total number of signs * 100%

5. CONCLUSION & FUTURE SCOPE

5.1. Conclusion

In this project work the area of hand-gesture recognition for applications in human-computer interaction is explored. The system developed here is a real time hand gesture recognition based power point slide control. The main problem of gesture recognition lies in the complexity of the classification algorithms, especially when using high dimensional feature vectors which become necessary in order to be able to distinguish several hundreds of gestures. Thus, the development of good classification methods & precise features is very important in order to run such systems in real-time. The Proposed approach is based on Zernike moments for feature extraction because ZM's are direction and scaling invariant transform and Neural Network for gesture recognition where the gesture images are trained and tested using the feed forward back propagation neural network. Therefore proposed approach got better results with 11 Zernike moments and achieved the recognition rate of the neural network up to 82.5 % for the gesture image with uniform background in light environment with minimum distance of 1 or 2 feet from the camera. The results show a significant accuracy in real time recognition.

5.2. Future Scope

The work presented in this project recognizes static gestures only. The work can be extended to be able to recognize dynamic gestures. The system deals with images with uniform background, but it can be made background independent. The network can be trained to the other types of images. It is important to consider increasing data size, so that it can have more accurate and highly performance system. The gestures can be used to control many different applications in order to make gesture recognition powerful and robust in area of HCI.

REFERENCES

[1] Vladimir I. Pavlovic, Rajeev, Sharma, and Thomas S. Huang, Fellow, "Visual Interpretation of Hand Gestures for Human-Computer Interaction: A Review", IEEE transactions on Pattern Analysis and Machine Intelligence, vol. 19, no. 7, July 1997.

[2] Hideyuki Sawada, Seiji Uta and Shojis Hashimoto, "Gesture Recognition for Human-Friendly Interface in Designer - Consumer Cooperate Design System", Proceedings of the 1999 IEEE International Workshop on Robot and Human Interaction Pisa, Italy - September 1999.

[3] Prateem Chakraborty, Prashant Sarawgi, Ankit Mehrotra, Gaurav Agarwal, Ratika Pradhan, "Hand Gesture Recognition: A Comparative Study", Proceedings of the International MultiConferenc.

[4] Elena Sánchez-Nielsen, "Hand Gesture Recognition for Human-Machine Interaction", Department of Statistic O.R. and Computer Science, University of La Laguna Edificio de Física y Matemáticas 38271, La Laguna, Spain.

[5] Pragati Garg, Naveen Aggarwal and Sanjeev Sofat, "Vision Based Hand Gesture Recognition", World Academy of Science, Engineering and Technology 49 2009.

[6] S.M. Hassan Ahmeda, Todd C. Alexanderb and Georgios C. Anagnostopoulosb, "Real-time, Static and Dynamic Hand Gesture Recognition for Human-Computer Interaction", Electrical Engineering, University of Miami. http://www.amalthea-reu.org/pubs/amalthea_tr_2008_01.pdf.

[7] Cristina Manresa, Javier Varona, Ramon Mas and Francisco J. Perales, "Real –Time Hand Tracking and Gesture Recognition for Human-Computer Interaction" Departamento de Matemáticas e Informática Universitat de les Illes Balears. WSEAS Transactions on Computers, Volume 9, Issue 6 (June 2010), Pages: 583-592, Year of Publication: 2010, ISSN: 1109-2750.

[8] Tushar Agrawal, Subhasis Chaudhuri, "Gesture Recognition using Position and Appearance features" Analog Devices Inc. Mumbai, India; IIT Bombay, Mumbai, India. http://www.ee.iitb.ac.in/~sc/papers/icip03-tush.pdf.

[9] Zhi Li, Ray Jarvis, "Real time Hand Gesture Recognition using a Range Camera", Australasian Conference on Robotics and Automation (ACRA), December 2-4, 2009, Sydney, Australia.

[10] G. R. S. Murthy & R. S. Jadon, "A review of Vision Based Hand Gestures Recognition", International Journal of Information Technology and Knowledge Management, July-December 2009, Volume 2, No. 2, pp. 405-410.

[11] Paulo Peixoto, Joao Carreira, "A Natural Hand Gesture Human Computer Interface using Contour Signatures", ISR - Institute of Systems and Robotics. http://www.isr.uc.pt/~joaoluis/papers/HCI2005-Final.pdf.

[12] Yikai Fang, Kongqiao Wang, Jian Cheng and Hanqing Lu, "A Real-Time Hand Gesture Recognition Method", National Lab of Pattern Recognition Institute of Automation, Chinese Academy of Sciences, Beijing 100080, China.

[13] Sushmita Mitra, Tinku Acharya, "Gesture Recognition: A Survey" IEEE Transactions on Systems, Man, and Cybernetics—Part C: Applications and Reviews, Vol. 37, No. 3, MAY 2007.

[14] Chin-Chen Chang, Jiann-Jone Chen, Wen-Kai Tai and Chin-Chuan Han," New Approach for Static Gesture Recognition", Journal of Information Science

[15] Rohit Verma, Ankit Dev, "Vision based Hand Gesture Recognition Using Finite State Machines and Fuzzy Logic", Ultra Modern

Telecommunications & Workshops, 2009. ICUMT '09. International Conference on 12-14 Oct. 2009.

[16] Matti-Antero Okkonen, Vili Kellokumpu, Matti Pietik¨ainen, and Janne Heikkil, “A Visual System for Hand Gesture Recognition in Human-Computer Interaction” Department of Electrical and Information Engineering. http://www.springerlink.com/index/g82687g1r2561j74.pdf.

[17] S. Mohamed Mansoor Roomi, R. Jyothi Priya and H. Jayalakshmi, “Hand Gesture Recognition for Human-Computer Interaction”,Department of Electronics and Communication, Thiagarajar College of Engineering, Madurai, Tamil Nadu, India, Journal of Computer Science 6, 2010 Science Publications.

[18] Lae Kyoung Lee , Kim, Young-Kiu Choani Man Hyung Lee, “Recognition of Hand gesture to Human-Computer Interaction” http://www.iconi.org/apic-ist/APIC-IST_ICONI%202009.pdf.

[19] Thiago R. Trigo and Sergio Roberto M. Pellegrino, “An Analysis of Features for Hand-Gesture Classification” IWSSIP 2010 - 17th International Conference on Systems, Signals and Image Processing Electronic Engineering and Computer Department Aeronautics Technological Institute (ITA), São José dos Campos, Brazil.

[20] Tin Hninn Hninn Maung, “Real-Time Hand Tracking and Gesture Recognition System Using Neural Networks”, World Academy of Science, Engineering and Technology 50 2009.

[21] Vaishali.S.Kulkarni, “Appearance Based Recognition of American Sign Language Using Gesture Segmentation”, (IJCSE) International Journal on Computer Science and Engineering, Vol. 02, No. 03, 2010, 560-565.

[22] N. Senthilkumaran1 and R. Rajesh, “Edge Detection Techniques for Image Segmentation – A Survey of Soft Computing Approaches”, International Journal of Recent Trends in Engineering, Vol. 1, No. 2, May 2009.

[23] Agn`es Just a Yann Rodriguez a S´ebastien Marcel a, “Hand Posture Classification and Recognition using the Modified Census Transform”. Published in Proceedings of the IEEE International Conference on Automatic Face and Gesture Recognition (AFGR), April 2006.

[24] Dr. Jane J. Stephan, Sana'a Khudayer, ”Gesture Recognition for Human-Computer Interaction (HCI)”, International Journal of Advancements in Computing Technology, Vol. 2, No. 4, pp. 30 ~ 35, 2010.

[25] Christos Stergiou and Dimitrios Siganos,”Neural Networks”. http://www.doc.ic.ac.uk/~nd/surprise_96/journal/vol4/cs11/report.html

[26] Olga miljkovic, “Image Pre-processing Tool”, kragujevac j. math, (2009). http://elib.mi.sanu.ac.rs/files/journals/kjm/32/kjom3209.pdf

[27] Christian von, Hardenberg François Bérard, ”Bare-Hand Human-Computer Interaction”, Proceedings of the ACM Workshop on Perceptive User Interfaces, Orlando, Florida, USA, Nov. 15-16 2001.

[28] Howard Demuth, Mark Beale, Martin Hagan, “Neural Network Toolbox™ 6”, User’s Guide

[29] Gholam Reza Amayeh, Ali Erol, George Bebis, and Mircea Nicolescu , ”Accurate and Efficient Computation of High Order Zernike Moments”

[30] Klimis Symeonidis, “Hand Gesture Recognition Using Neural Networks, School of Electronic and Electrical Engineering
On August 23, 2000
http://personal.ee.surrey.ac.uk/Personal/T.Windeatt/msc_projects/symeonidis/Final%20Project.pdf