International Journal of Modern Engineering Research (IJMER)
www.ijmer.com Vol.2, Issue.4, July-Aug 2012 pp-2857-2861 ISSN: 2249-6645

Performance Comparison of Three Different Classifiers for Hci Using Hand Gestures

Anand.H.Kulkarni[1], Sachin.A.Urabinahatti[2]
Department of Computer Science and Engineering
Gogte Institute Of Technology, Belgaum, Karnataka, India

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 paper, a robust hand gesture recognition system is presented for recognizing static gestures based on Zernike moments (ZMs) using Three Classifiers K-nearest Neighbour(KNN), Support Vector Machine (SVM) and Artificial Neural Network (ANN) . 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, and a comparative study is carried out to show which classifier works better in reorganisation of gestures.*

Keywords: *Hand Gesture Recognition, Human Computer Interaction, KNN, SVM, ANN, Zernike Moments.*

I. INTRODUCTION

The way humans interact with computers is constantly evolving, with the general purpose of 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 fingers, hands, arms, head, face, or body with the intent of:

1) Conveying meaningful information or
2) Interacting with the environment [2].

Gestures can be static or dynamic some gestures also have both static and dynamic elements, as in sign languages. A hand gesture is defined as a dynamic movement referring to a sequence of hand postures connected by continuous motions over a short time span, such as waving good-bye; hand posture is a static hand pose without involvement of movements. Hand gesture recognition finds applications in varied domains including virtual environments, smart surveillance, sign language translation, medical systems etc. 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. In Human Computer interaction, to operate computer with hand gestures no accessories like gloves are needed. The approaches to implement Human-Computer Interaction using Hand Gestures are based on k-Nearest Neighbour (KNN), Support Vector Machines (SVM) and Artificial Neural Network (ANN) Methods.

II. METHODOLGY

The methodological analysis of the present work has been presented pictorially in Figure 1. The work commence with capturing Hand gesture images using cameras or scanners. These images are made to undergo pre-processing steps like filtering and segmentation. Then different texture and colour features are extracted from the processed image. Finally, the feature values are fed as input to the classifier (KNN, ANN, SVM) to classify the given image.

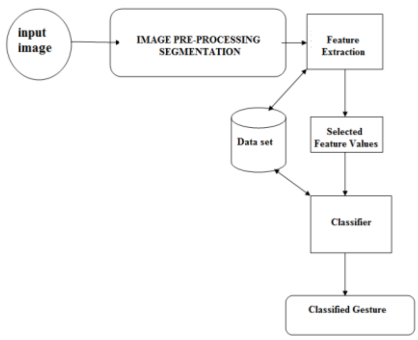


Fig 1: Block Diagram

A. **Input Image**: The first step in the proposed approach is to capture the video from the digital camera and extract the video frames. The video is captured from the digital camera and the captured frames are then stored in the image database.

B. **Image Database**: The next point in the project is creation of the image database with all the images 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 200 image samples; where 40 image samples for each gesture 1, gesture 2, gesture 3, gesture 4 and gesture 5. The image database itself is responsible for the better efficiency of the classifier as it is that which decides the robustness of the algorithm.

Gesture 1 Gesture 2 Gesture 3 Gesture 4 Gesture 5

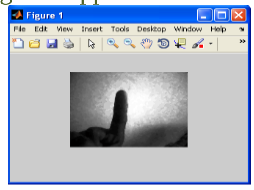
C. **Image Pre-processing:** Image pre-processing is the name for operations on images at the lowest level of abstraction whose aim is an improvement of the image data that suppress undesired distortions or enhances some image features important for further processing and analysis task. It does not increase image information content. Its methods use the considerable redundancy in images. Neighbouring pixels corresponding to one real object have the same or similar brightness value. If a distorted pixel can be picked out from the image, it can be restored as an average value of neighbouring pixels. In the proposed approach image pre-processing methods are applied to the frames extracted from the video captured and which are stored in image database. The image processing method i.e., median filtering is used in order to remove noise and preserve the edges in the image, Edge detection for segmentation i.e., to extract the region of interest (hand part) from the image.

i. Median Filtering

The median filter is a nonlinear digital filtering technique, often used to remove noise. Such noise reduction is a typical pre-processing step to improve the results of later processing (for example, edge detection on an image). Median filtering is a nonlinear operation often used in image processing to reduce "salt and pepper" noise. Median filtering is a non-linear smoothing method that reduces the blurring of edges and significantly eliminates impulse noise. It suppresses image noise without reducing the image sharpness and can be applied iteratively.

ii. *Segmentation*

Image segmentation is process i.e., used to simplify and/or change the representation of an image into something that is more meaningful and easier to analyze. As the premise of feature extraction and pattern recognition, image segmentation is one of the fundamental approaches of digital image processing. Image Segmentation is the process that is used to distinguish object of interest from background. The proposed approach uses edge as a factor to extract the region of interest from the given image. Edges are pixels where the intensity image function changes abruptly. Edge detectors are collection of local image pre-processing methods used to locate changes in the brightness function. Edges are the sign of lack of continuity, and ending. EDGE function finds edges in intensity image. EDGE function takes image as its input, and returns a binary image of the same size, with 1's where the function finds edges in image and 0's elsewhere. The figure 2 shows the filtered image and the edges in the segmented image that is detected using EDGE function.



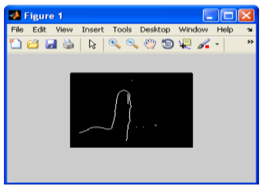


Fig 2a: Before segmentation Fig 2b: After segmentation

D. 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. 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. The proposed approach considers 11 order of moments to extract the feature from the segmented image.

- ***Zernike moments (ZM)***

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 we adopt 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:

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

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 $|m|\le n,\ge x2+y2$

Θ is the angle between vector ρ and the *x* -axis in a counter clockwise direction.

The form of orthogonal Zernike basis polynomials, $Vnm(\rho, \theta)$

$Vnm(\rho, \theta)$ is a complex conjugate of $Vnm(\rho, \theta)$ = $R_{nm}(\rho)e^{-jm\Theta}$ 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:

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

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 [10].

D. Gesture Recognition & Classification:

The recognition process consists of two phases, training and classification. Classification of gestures is done using KNN (k-Nearest Neighbor), ANN (Artificial Neural Network) and SVM (Support Vector Machine) classifiers.

i. K-Nearest Neighbour

KNN stands for "k-nearest neighbour algorithm", it is one of the simplest but widely used machine learning algorithm. An object is classified by the "distance" from its neighbours, with the object being assigned to the class most common among its k distance-nearest neighbours. If k = 1, the algorithm simply becomes nearest neighbour algorithm and the object is classified to the class of its nearest neighbour [11][12].

ii. Artificial Neural Network

An Artificial Neural Network (ANN) is an information processing paradigm that is inspired by the way biological nervous systems, such as the brain, process the information. The key element of this paradigm is the novel structure of the information processing system. It is composed of a large number of highly interconnected processing elements (neurons) working in unison to solve specific problems. ANNs, like people, learn by example. An ANN is configured for a specific application, such as pattern recognition or data classification, through a learning process. A trained neural network can be thought of as an "expert" in the category of information it has been given to analyze [9][13].

iii. Support Vector Machine

SVMs are currently a hot topic in the machine learning community, creating a similar enthusiasm at the moment as Artificial Neural Networks used to do before. Far being, SVMs yet represent a powerful technique for general (nonlinear) classification, regression and outlier detection with an intuitive model representation. Support vector machines are a set of related supervised learning methods used for classification and regression. Given a set of training examples, each marked as belonging to one of two categories, an SVM training algorithm builds a model that predicts whether a new example falls into one category or the other [14][15][16].

III. Expiremental Analysis & Results

- *Experimental Analysis WRT Number of features v/s KNN efficiency*

Number of Features	Recognition Rate for Gesture 1 (%)	Recognition Rate for Gesture 2 (%)	Recognition Rate for Gesture 3 (%)	Recognition Rate for Gesture 4 (%)	Recognition Rate for Gesture 5 (%)	Overall KNN Efficiency (%)
2	65	45	58	56	51	55
3	82	70	78	74	76	76
4	77	51	58	62	72	64
5	56	48	49	56	51	52
6	70	55	60	61.5	66	62.5
7	85	70	72.5	78	82	77.5
8	79	59	65	67	75	69
9	79	61	68	65	74.5	69.5
10	71	73	74	71.5	70.5	72
11	85	70	79	76	77.5	77.5

Table 1: Number of features v/s KNN efficiency.

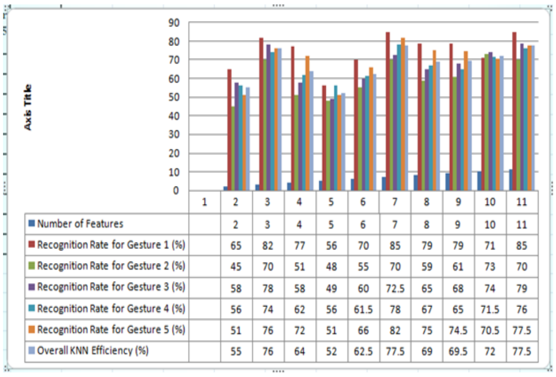


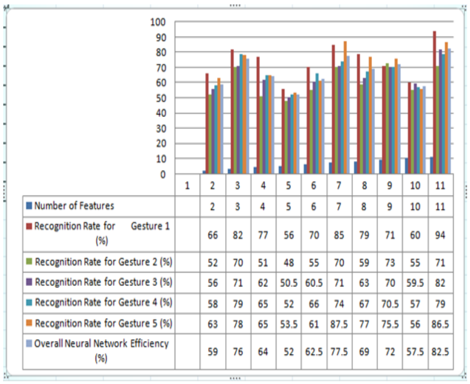
Figure 3: Graphical Analysis for Number of Features v/s KNN efficiency

The table 1 shows the dependency of the KNN 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 KNN is optimum when there are 11 order of moments i.e., 11 feature vector. The above figure 3. Shows the graphical representation of analysis with respect to Number of Features v/s KNN Efficiency which shows the system is optimum when there are 11 order of Zernike moments.

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

Number of Features	Recognition Rate for Gesture 1 (%)	Recognition Rate for Gesture 2 (%)	Recognition Rate for Gesture 3 (%)	Recognition Rate for Gesture 4 (%)	Recognition Rate for Gesture 5 (%)	Overall Neural Network Efficiency (%)
2	66	52	56	58	63	59
3	82	70	71	79	78	76
4	77	51	62	65	65	64
5	56	48	50.5	52	53.5	52
6	70	55	60.5	66	61	62.5
7	85	70	71	74	87.5	77.5
8	79	59	63	67	77	69
9	71	73	70	70.5	75.5	72
10	60	55	59.5	57	56	57.5
11	94	71	82	79	86.5	82.5

Table 2: Number of features v/s Neural Network efficiency



	1	2	3	4	5	6	7	8	9	10	11
Number of Features		2	3	4	5	6	7	8	9	10	11
Recognition Rate for Gesture 1 (%)		66	82	77	56	70	85	79	71	60	94
Recognition Rate for Gesture 2 (%)		52	70	51	48	55	70	59	73	55	71
Recognition Rate for Gesture 3 (%)		56	71	62	50.5	60.5	71	63	70	59.5	82
Recognition Rate for Gesture 4 (%)		58	79	65	52	66	74	67	70.5	57	79
Recognition Rate for Gesture 5 (%)		63	78	65	53.5	61	87.5	77	75.5	56	86.5
Overall Neural Network Efficiency (%)		59	76	64	52	62.5	77.5	69	72	57.5	82.5

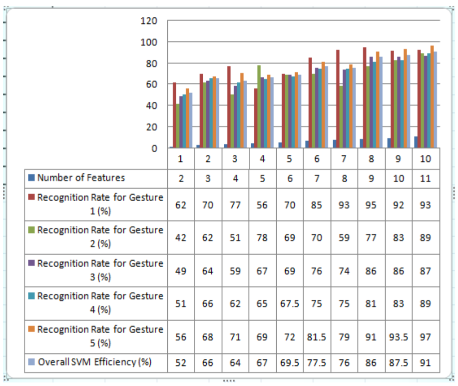
Figure 4: Graphical Analysis for Number of Features v/s ANN efficiency

The table 2 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 4. 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.

- *Experimental Analysis WRT Number of features v/s SVM efficiency*

Number of Features	Recognition Rate for Gesture 1 (%)	Recognition Rate for Gesture 2 (%)	Recognition Rate for Gesture 3 (%)	Recognition Rate for Gesture 4 (%)	Recognition Rate for Gesture 5 (%)	Overall SVM Efficiency (%)
2	62	42	49	51	56	52
3	70	62	64	66	68	66
4	77	51	59	62	71	64
5	56	78	67	65	69	67
6	70	69	69	67.5	72	69.5
7	85	70	76	75	81.5	77.5
8	93	59	74	75	79	76
9	95	77	86	81	91	86
10	92	83	86	83	93.5	87.5
11	93	89	87	89	97	91

Table 3: Number of features v/s SVM efficiency



	1	2	3	4	5	6	7	8	9	10
Number of Features	2	3	4	5	6	7	8	9	10	11
Recognition Rate for Gesture 1 (%)	62	70	77	56	70	85	93	95	92	93
Recognition Rate for Gesture 2 (%)	42	62	51	78	69	70	59	77	83	89
Recognition Rate for Gesture 3 (%)	49	64	59	67	69	76	74	86	86	87
Recognition Rate for Gesture 4 (%)	51	66	62	65	67.5	75	75	81	83	89
Recognition Rate for Gesture 5 (%)	56	68	71	69	72	81.5	79	91	93.5	97
Overall SVM Efficiency (%)	52	66	64	67	69.5	77.5	76	86	87.5	91

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

The table 3 shows the dependency of the SVM 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 SVM is optimum when there are 11 order of moments i.e., 11 feature vector. The above figure 5.Shows the graphical representation of analysis with respect to Number of Features v/s KNN Efficiency which shows the neural network is optimum when there are 11 order of Zernike moments

- *Comparative result of three classifiers*

Classifier	System efficiency
KNN	77.5%
ANN	82.5%
SVM	91%

Table 4: System Efficiency Using Three Different Classifiers

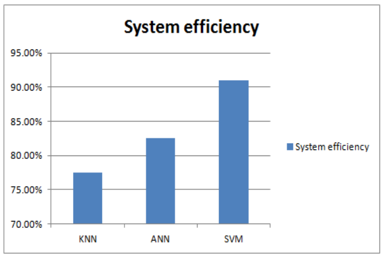


Figure 6: Graphical Analysis of System Efficiency Using Three Different Classifiers

IV. CONCLUSION

In this project work the area of hand-gesture recognition is introduced. The system developed here is a real time hand gesture recognition, 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. Therefore proposed approach got better results with 11 Zernike moments and achieved the recognition rate of the KNN up to 77.5%, ANN up to 82.5% and SVM up to 91% for the gesture image with uniform background in light environment with minimum distance of 1 or 2 feet. The results show a significant accuracy in real time recognition. So by this we can conclude this system works with better efficiency considering SVM as a classifier either then ANN or KNN classifier.

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] 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.

[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.

[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] 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.

[9] Christos Stergiou and Dimitrios Siganos , "NeuralNetworks".

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

[11] Image Classification Using SVM, KNN and Performance Comparison with Logistic Regression Team member: Qi Gu, Zhifei Song

[12] http://en.wikipedia.org/wiki/K-nearest_neighbor_algorithm

[13] Klimis Symeonidis, "Hand Gesture Recognition Using Neural Networks, School of Electronic and Electrical Engineering

[14] R. Gonzalez and R. Woods, "Digital image processing", 3rd Edition 2008, Prentice Hall, New York.

[15] Sukhwinder Singh, Vinod Kumar, *Senior Member, IEEE*, H.K. Verma, and Dilbag Singh, "SVM Based System for classification of Microcalcifications in Digital Mammograms", Proceedings of the 28th IEEE EMBS Annual International Conference New York City, USA, Aug 30-Sept 3, 2006.

[16] Faramarz Dehghan, Hamid Abrishami-Moghaddam," Comparison Of SVM And Neural Network Classifiers In Automatic Detection Of Clustered Microcalcifications In Digitized Mammograms", Proceedings Of The Seventh International Conference On Machine Learning And Cybernetics, Kunming, 12-15 July, IEEE 2008.