

# A Neural Network based Real Time Hand Gesture Recognition System

Tasnuva Ahmed  
Lecturer, Dept. of CSIT  
Southern University Bangladesh

## ABSTRACT

Hand Gesture is habitually used in every day life style. It is so natural way to communicate. Hand gesture recognition method is widely used in the application area of Controlling mouse and/or keyboard functionality, mechanical system, 3D World, Manipulate virtual objects, Navigate in a Virtual Environment, Human/Robot Manipulation and Instruction Communicate at a distance. This paper introduces a real time hand gesture recognition system. This system consists of three stages: image acquisition, feature extraction, and recognition. In the first stage input image of hand gestures are acquiesced by digital camera in approximate frame rate. In second stage a rotation, translation, scaling and orientation invariant feature extraction method has been introduce to extract the feature of the input image based on moment feature extraction method. Finally, a neural network is used to recognize the hand gestures. The performance of the system tested on real data. Based on the experimental results, we noted that this system shows satisfactory performance in hand gesture recognition.

## General Terms

Artificial Neural Network, Image Processing.

## Keywords

Hand gesture, Moment feature extraction method, Multilayer feed forward neural network.

## 1. INTRODUCTION

Interaction between human and computer has become an important part in everyday life. It concern with the design, evaluation and implementation of interactive computing systems for human use [1]. The most popular communicating devices are keyboard, mice, light pen, trackball, keypads etc. The devices are familiar but not so natural for communication. To the development of vision-based interface, now computer can able to see which make the human computer interaction rich. These employ a new communication and controlling facility that is more reliable, user-friendly and time saving. A number of research work have done on hand gesture recognition based on different methods and techniques. Among of them the Hand Gesture Recognition Using Computer Vision developed by Ray Lockton [2] in which he build a man-machine interface using a video camera to interpret the American one-handed sign language alphabet and number gestures. Klimis Symeonidis developed Hand Gesture Recognition Using Neural Networks [3] which used orientation histogram to recognize static hand gestures. Hand Gesture Recognition Based on Decision Tree developed by

Yiqiang CHEN, Wen GAO, Jiyong MA [4] present a new technique for the recognition of hand gesture using decision tree method based on information entropy. Hand-Gesture Recognition for HCI developed by Atid Shamaie, Wu Hai and Alistair Sutherland [5] is a system to recognize hand gestures from real-time video images. Hand Gesture Recognition using Input-Output Hidden Markov Models developed by Sebastien Marcel, Olivier Bernier, Jean-Emmanuel Viallet and Daniel Collobert [6] this method deals with the dynamic aspects of gestures. Hand Gesture Recognition in Camera-Projector System developed by Attila Licsár, Tamás Szirányi [7] proposes a vision-based hand gesture recognition system. Vision-Based Recognition of Continuous Dynamic Hand Gestures developed by Yuanxin Zhu [8] A spatio-temporal appearance model is proposed to represent dynamic hand gestures. Hand gesture recognition for Human Machine Interaction developed by Elena Sánchez-Nielsen, Luis Antón-Canalís, Mario Hernández-Tejera [9] paper is a real time vision system for its application within visual interaction environments through hand gesture recognition, using general-purpose hardware and low cost sensors. Real-time gesture recognition using deterministic boosting developed by Raymond Lockton and Andrew W. Fitzgibbon [10] a gesture recognition system which can reliably recognize single-hand gestures in real time. Camera-based Gesture Recognition for Robot Control developed by Andrea Corradini, Horst-Michael Gross [11]. In these systems the recognition engine is mainly based on three algorithms: dynamic pattern matching, statistical classification, and neural networks. The model-based static gesture recognition approach pro-posed by Davis and Shah [12], use a finite-state machine to model four qualitatively distinct phases of a generic gesture. A dynamic gesture recognition system for American Sign Language (ASL) interpretation has been developed by Charayaphan et al. [13] to detect the direction of hand motion by tracking the hand location.

The objective of this paper is to design a hand gesture recognition system that works in real-time and recognized manipulative hand gestures. The gestures that are used in this recognition system have distinct meaning. Each one of these gestures represents a particular job. This paper concern to design a feature extraction method that is invariant over rotation, scaling, translation and orientation based on moment feature extraction method [14] so that the system can recognize hand gestures captured in different angle or orientation or size.

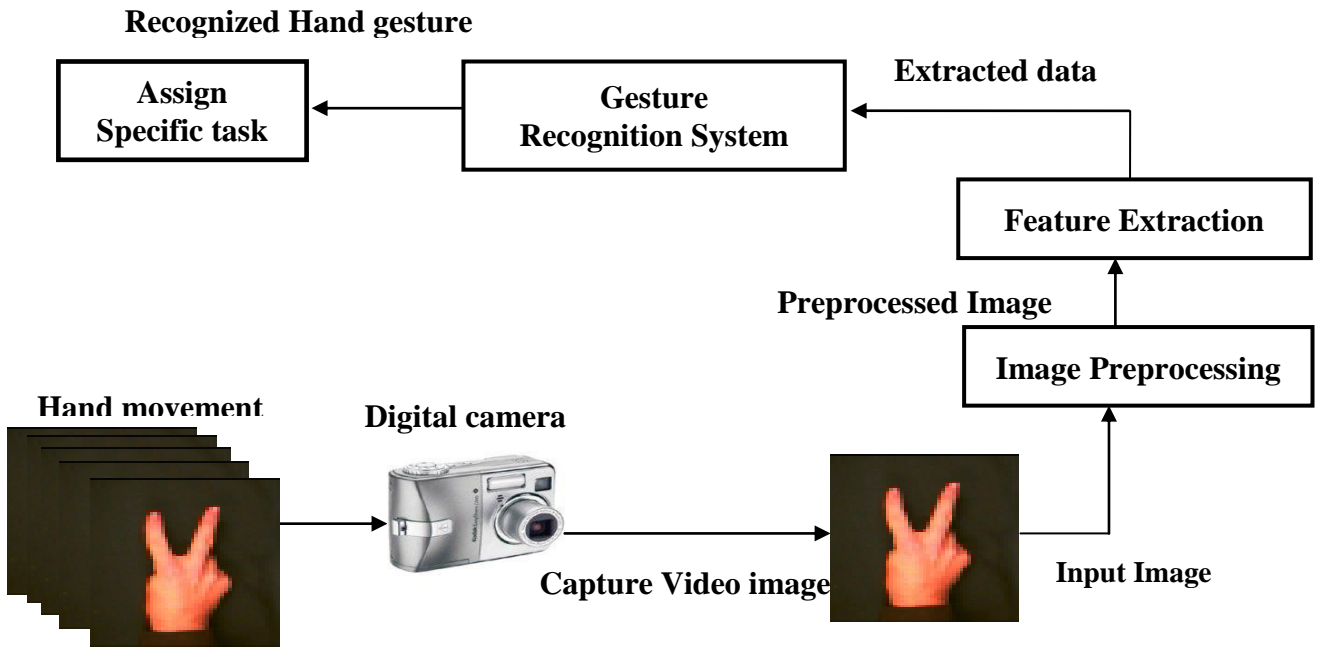


Fig1: Architecture of the hand gesture recognition system

## 2. SYSTEM ARCHITECTURE

The system proposed in this paper is designed to recognize hand gesture in real-time. The technique that is used to recognize hand gesture is based on computer vision. The overall system architecture is shown in figure 1. The whole system of hand gesture recognition is divided into four phases: Image Acquisition, Image Preprocessing, Feature Extraction and Hand Gesture Recognition.

## 3. IMAGE ACQUISITION

The user simply is moving his/her hand into view of the camera to begin interaction with the system. The system uses a *kodak-easyshare-c340* digital camera to capture the desk area where the user's hand is located. The background of the image is black and has good lighting system. Image is acquired by approximate frame rate. To find the change of hand pose, compare two frames using a threshold value. The threshold value is 30. At first two consecutive frames are subtracted and if the subtracted result is greater than the threshold values (i.e.  $T=30\%$ ), then the next frame is considered as different hand pose. Otherwise they are assumed to be the same hand pose. If  $f(x, y, n)$  and  $f(x, y, n-1)$  are two consecutive frames and  $D(x, y, n)$  is the difference of those two consecutive frames obtained by following equation:

$$D(x, y, n) = |f(x, y, n) - f(x, y, n-1)| \dots \dots \dots (1)$$

If  $D(x, y, n) > T$  capture new frame otherwise keep the count frame. Where  $n$  is the number of frames and  $(x, y)$  is the coordinate of the frame.

## 4. IMAGE PREPROCESSING

Images that are acquired from video sequence, passed through the preprocessing phase to obtain enriched data for feature extraction.

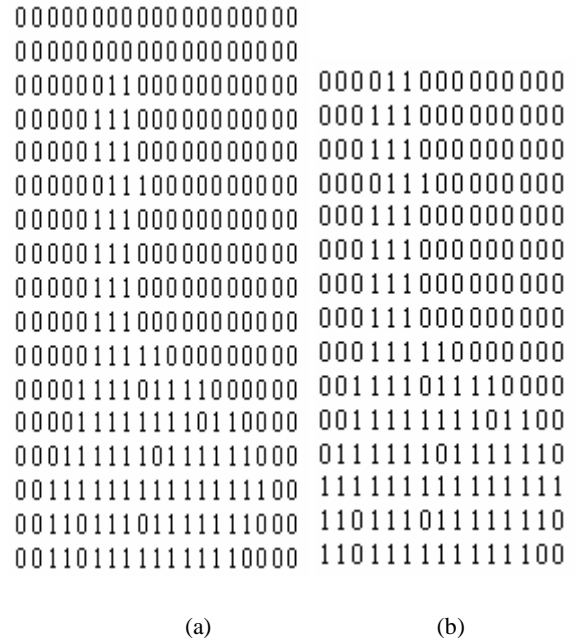


Fig 2: (a) Original image (b) Reduced image.

The algorithms sequentially used in this phase are graying, normalizing and histogram equalizing. After obtaining a grayscale image, we use Median filters to remove noise from the input image, and then threshold the image to a binary one. We used the local adaptive threshold algorithm for the binarization step. Then remove the unwanted portion of the image; we have to remove the unnecessary pixels (0) from

original image. This is done because we need to develop size independent algorithm.

**Algorithm 1**

- 1) Start from top-left corner; repeat for each column and row.  
If sum of all black pixels in row/column > 0  
Then save column and row
- 2) Else don't save column/row.

**5. FEATURE EXTRACTION**

To extract the feature of the hand gesture we developed a rotation, scaling, translation and orientation independent feature extraction method and obtain 33 features for hand gesture.

**Center of the image**

Center of the image can obtain by using following equation:

$$Center\_x = width / 2 \dots \dots \dots (4)$$

$$Center\_y = height / 2 \dots \dots \dots (5)$$

**Feature 1**

The first feature is the relation between the height and the width of the hand gesture

$$feature\ 1 = height / width \dots \dots \dots (6)$$

**Feature 2 - 25**

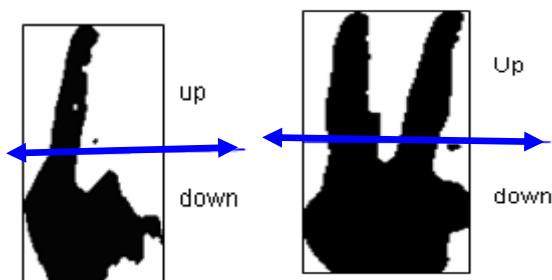
These features check how the black pixels are distributed in the image. First the number of pixels inside the image is calculated that is *total\_pixels* of hand gestures.

$$Total\_pixel = height \times weight \dots \dots \dots (7)$$

The feature 2 and 3 are the percentage of black pixels located in the upper and lower areas of hand gestures, in other words, the pixels located up and down the central point.

$$feature2 = up\_pixels / total\_pixels \dots \dots \dots (8)$$

$$feature3 = down\_pixels / total\_pixels \dots \dots \dots (9)$$

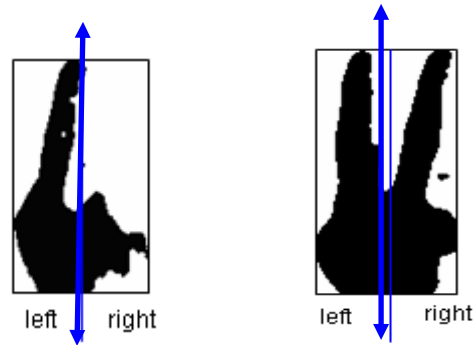


**Fig 3: divide the image into upper and lower area.**

The feature 4 and 5 are the percentage of black pixels located in the left and right areas of hand gestures, in other words, the pixels located in the left and right of the central point.

$$feature4 = left\_pixels / total\_pixels \dots \dots \dots (10)$$

$$feature5 = right\_pixels / total\_pixels \dots \dots \dots (11)$$

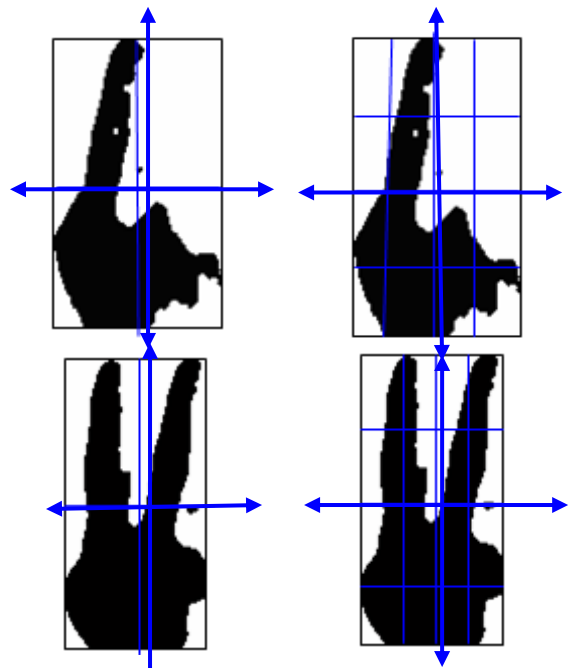


**Fig 4: divide images into left and right areas.**

Now split the image into 4 sub regions and calculate the percentage of black pixels located in every region. Again sub divides every region into four and calculates the percentage of black pixels of those regions. The features 6 to 9 are the percentage of black pixels located in the every sub areas of hand gestures.

$$feature_n = sub\_area\_pixels_n / total\_pixels \dots \dots (12)$$

Where n = 6 to 25



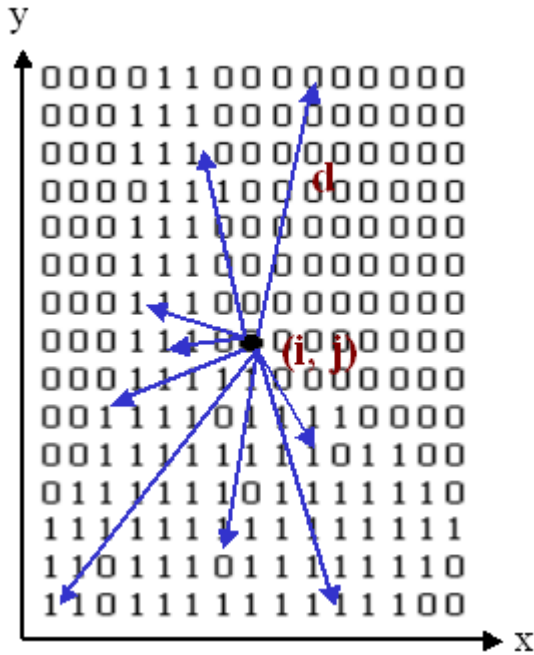
**Fig 5: Split the image region**

**Feature 26**

The feature 26 calculate average distance between all the black pixels and the central point that invariant object rotation.

$$feature\ 26 = \frac{1}{Total\_Pixels} \times \sum_y \sum_x \sqrt{(x-i)^2 \times (y-j)^2} \dots \dots \dots (13)$$

Where (x, y) are the coordinates of a pixel and (i, j) are the coordinates of central point.



**Fig 6: Calculate distance between black pixels and central point.**

**Feature 27-33**

Calculate the central moments of hand gesture. For  $f(x, y)$  2-dimensional function of  $M \times N$  binary image, the moment of order  $(p + q)$  is defined by:

$$m_{pq} = \sum_{x=1}^M \sum_{y=1}^N (x)^p (y)^q f(x, y) \dots\dots\dots(14)$$

Where  $p, q = 0, 1, 2, 3 \dots\dots\dots$

Central moment obtains by following equation:

$$\mu_{pq} = \sum_x \sum_y (x - \bar{x})^p (y - \bar{y})^q f(x, y) \dots(15)$$

Where  $\bar{x} = \frac{m_{10}}{m_{00}}$  and  $\bar{y} = \frac{m_{01}}{m_{00}}$

For scaling normalization the central moment changes as following equation:

$$\eta_{pq} = \frac{\mu_{pq}}{\mu_{00}^\gamma} \dots\dots\dots(16)$$

where  $\gamma = \left[ \frac{(p + q)}{2} \right] + 1$

seven values, computed by normalizing central moments through order three, that are invariant to object scale, position, and orientation. In terms of the central moments, the seven moments are given as,

$$M_1 = \eta_{20} + \eta_{02} \dots\dots\dots(17)$$

$$M_2 = (\eta_{20} - \eta_{02})^2 + 4\eta_{11}^2 \dots\dots\dots(18)$$

$$M_3 = (\eta_{30} - 3\eta_{12})^2 + (3\eta_{21} - \eta_{03})^2 \dots\dots(19)$$

$$M_4 = (\eta_{30} + \eta_{12})^2 + (\eta_{21} + \eta_{03})^2 \dots\dots(20)$$

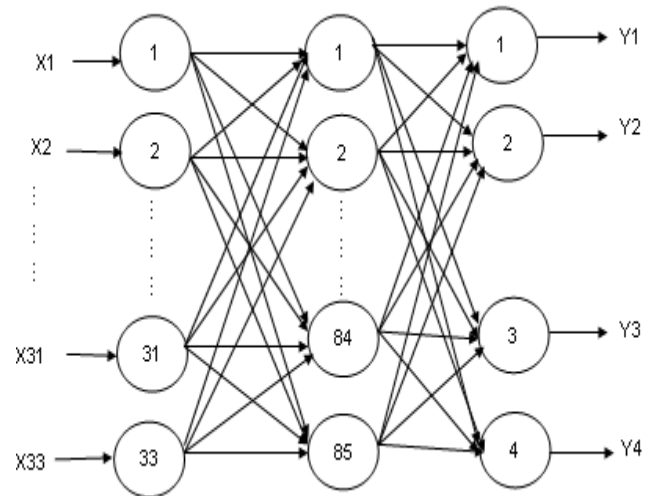
$$M_5 = (\eta_{30} - 3\eta_{12})(\eta_{30} + \eta_{12})[(\eta_{30} + \eta_{12})^2 - 3(\eta_{21} + \eta_{03})^2] + (3\eta_{21} - \eta_{03})(\eta_{21} + \eta_{03})[3(\eta_{30} + \eta_{12})^2 - (\eta_{21} + \eta_{03})^2] \dots\dots\dots(21)$$

$$M_6 = (\eta_{20} - \eta_{02})[(\eta_{30} + \eta_{12})^2 - (\eta_{21} + \eta_{03})^2] \dots + 4\eta_{11}(\eta_{30} + \eta_{12})(\eta_{21} + \eta_{03}) \dots\dots\dots(22)$$

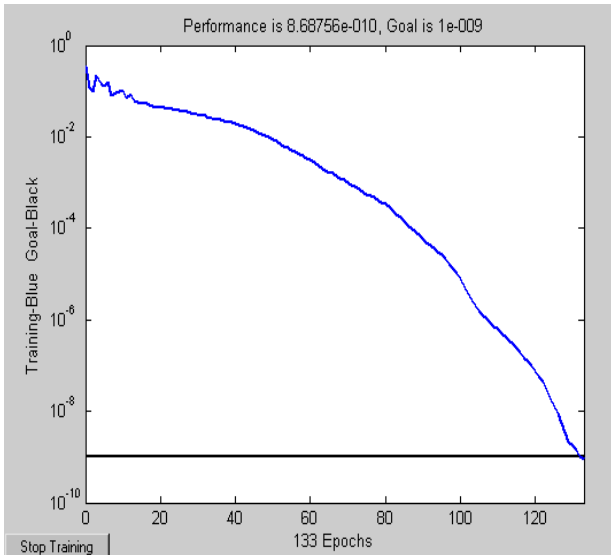
$$M_7 = (3\eta_{21} - \eta_{03})(\eta_{30} + \eta_{12})[(\eta_{30} + \eta_{12})^2 - 3(\eta_{21} + \eta_{03})^2] - (\eta_{30} + 3\eta_{12})(\eta_{21} + \eta_{03})[3(\eta_{30} + \eta_{12})^2 - (\eta_{21} + \eta_{03})^2] \dots\dots\dots(23)$$

**6. HAND GESTURE RECOGNITION**

We used a feed-forward multilayer ANNs. The architecture of the network is 33-85-4 shown in figure 7. For this network, we used 85 hidden neurons by calculating the input vector dimension that is  $33 \times 20$  and the output layer has 4 neurons because there are 4 outputs of the network. The network is fully connected and use back-propagation learning algorithm [15].



**Fig7: Network design for the system**



**Fig 8: Training of the Neural Network**

Train applies the inputs to the new network, calculates the outputs, compares them to the associated targets, and calculates a mean square error. If the error goal is met, or if the maximum number of epochs is reached, the training is stopped and train returns the new network and a training record. Otherwise train goes through another epoch. The figure 8 shows the training of the network. At the beginning the Sum squared error (sse) is big, but after 100 epochs the error of the network is very small and the neuronal net can recognize the training data with a very small error.

## 7. EXPERIMENTAL ANALYSIS AND RESULT





To developed a real-time hand gesture recognition system we use MATLAB that is a high performance language for computing. To get efficient response from this network we use 5 samples for each type of hand gesture. Instead we present to the network 33 element vectors, each one representing a single image. A dataset of 20 images after feature extraction contains  $33 \times 20$  elements. We have tested the recognition method with 4 gesture types. In Table 1 the recognition efficiency of hand gesture classification method with several trainer and tester users can be seen. The trainer users can be found at the columns and the tester users are at the rows. It is tested by a set of 50 samples per user.

**Table 1: Recognition result**

test set	Recognition Results				
	Train set				
	Set1	Set2	Set3	Set4	Set5
Set1	99.8%	96.1%	92.7%	86.7%	90.5%
Set2	94.5%	98.7%	90.3%	88.4%	82.5%
Set3	94.5%	90.3%	99.8%	96.1%	92.7%
Set4	90.3%	88.4%	82.5%	99.1%	94.5%
Set5	86.7%	92.7%	96.1%	94.5%	99.5%

If the train and test set is same the recognition rates are above 97%, otherwise the results are above 86 %. The overall recognition rates of test data are given bellow.

**Table 2: Overall recognition rate of each type of hand gesture**

Type of Pattern	No. of frames	Correct recognition rate	Error rate
	100	97%	3%
	100	86.5%	13.5%
	100	79.7%	30.3%
	100	91.6%	8.5%
Total	400	<b>88.7%</b>	13.8%

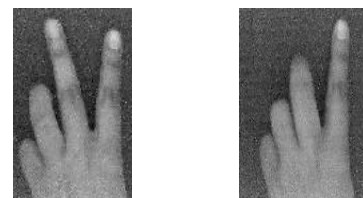
We use 100 frames for each type of hand gesture and evaluate correct recognition rate and the error rate of the system. We use the following equation to find the correct recognition rate:

$$\text{correct\_recognition\%} = \frac{\text{correct\_recognition}}{\text{total\_frames}} \quad (24)$$

And fin the error rate by using following equation:

$$\text{Error rate} = \frac{\text{false\_recognition}}{\text{total\_frames}} \quad (25)$$

Hence it would be more efficient if the system could learn only the faulty detected gestures by interaction with the user. The proposed system runs in simultaneous real-time at resolution of  $384 \times 288$  pixels on a single 1.7GHz Pentium processor. Sometime default frames are captured by the system. In this case the system cannot recognize hand gestures. The default frames are as shown in figure 9.



**Fig 9: default frames captured during acquisition**

## 8. CONCLUSION

The proposed approach applies computer vision methodology that is flexible, sound act in real time performance. This paper describes a new method to extract gesture features, which make the system rotation, scaling and translation independent. The gesture recognition system recognizes 88.7% hand gestures among the acquiesced frames. In the feature, the

system focuses on the three areas. They are-the system can detect and extract human hand from complex image that is an image where a human body is appeared, modify the system so that it can work in any lighting condition and expand the system to recognize the hand tracking.

## 9. REFERENCES

- [1] Alan Dix, Janet Finlay, Gregory D.Abowd, and Russell Beale, "Human-Computer Interaction" 3rd ed. Prentice Hall, 2003.
- [2] Ray Lockton, "Hand Gesture Recognition Using Computer Vision," 4<sup>th</sup> year project report, Balliol College, Oxford University.
- [3] Klimis Symeonidis, "Hand Gesture Recognition Using Neural Networks," Degree of Master of Science in Multimedia Signal Processing communications, School of Electronic and Electrical Engineering, On August 23, 2000.
- [4] Yiqiang CHEN, Wen GAO, Jiyong MA, "Hand Gesture Recognition Based on Decision Tree," Institute of Computing Technology, Chinese Academy of Sciences, Beijing.
- [5] Atid Shamaie, Wu Hai and Alistair Sutherland, "Hand-Gesture Recognition for HCI," *ERCIM News*, No.46, July 2001. [Online]. Available: [www.ercim.org/publication/Ercim\\_News/enw\\_46/shamaie.html](http://www.ercim.org/publication/Ercim_News/enw_46/shamaie.html). [Accessed August 3, 2006].
- [6] Sebastien Marcel, Olivier Bernier, Jean-Emmanuel Viallet and Daniel Collobert, "Hand Gesture Recognition using Input-Output Hidden Markov Models," France Telecom CNET, 2 avenue Pierre Marzin, 22307 Lannion, FRANCE.
- [7] Attila Licsár<sup>1</sup>, Tamás Szirányi<sup>1,2</sup>, "Hand Gesture Recognition in Camera-Projector System," <sup>1</sup>University of Veszprém, Department of Image Processing and Neurocomputing, H-8200 Veszprém, Egyetem u. 10. Hungary, <sup>2</sup>Analogical & Neural Computing Laboratory, Computer & Automation Research Institute, Hungarian Academy of Sciences, H-1111 Budapest, Kende u. 13-17, Hungary.
- [8] Yuanxin Zhu, "Vision-Based Recognition of Continuous Dynamic Hand Gestures," Department of Computer Science & Technology, Tsinghua University, Beijing, China.
- [9] Elena Sánchez-Nielsen, Luis Antón-Canalís, Mario Hernández-Tejera, "Hand gesture recognition for Human Machine Interaction," Department of Statistic,O.R. and Computer Science, University of La Laguna Edificio de Físicay Matemáticas, 38271, La Laguna, Spain.
- [10] Raymond Lockton and Andrew W. Fitzgibbon, "Real-time gesture recognition using deterministic boosting," Department of Engineering Science, University of Oxford, BMVC 2002.
- [11] Andrea Corradini, Horst-Michael Gross, "Camera-based Gesture Recognition for Robot Control," Technical University of Ilmenau, Department of Neuroinformatics, D-98684 Ilmenau, Federal Republic of Germany, IJCNN 2000.
- [12] Hunter E, Schlenzig J, Jain R, "Posture Estimation in Reduced-Model Gesture Input System," *Proc. of Int. Workshop on Automatic Face and Gesture Recognition*, 1995, Zurich, Switzerland.
- [13] Triesch J, von der Malsburg C, "Robost Classification of HandPosture againt Complex Background," *Proc. of Int. Workshop on Automatic Face and Gesture Recognition*, Vermont, Oct. 14-16 1996, pp. 170-175.
- [14] M. K. Hu, "Visual Pattern Recognition by Moment Invariants", *IRE Trans. Info. Theory*, vol. IT-8, pp.179-187, 1962.
- [15] Simon Haykin, "*Neural Networks: A Comprehensive Foundation*", Second edition, Pearson Education Asia, 2001.