HAND GESTURE CONTROLLED ROBOT WITH FACE RECOGNITION AND DETECTION CAPABILITIES

Rashmi Vashisth[1], Akshit Sharma[2], Shantanu Malhotra[3], Saurabh Deswal[4], Aman Budhraja[5]
Department of Electronics & Communication
Amity School of Engineering & Technology
New Delhi, India
Email: rashmiapj@gmail.com[1], akshits030@gmail.com[2], shantanu0507@gmail.com[3], deswal.saurabh94@gmail.com[4], amanbudhraja06@gmail.com[5]

Abstract: - In this paper, we introduced an accelerometer (ADXL335) based Gesture controlled Robot with ATmega16 microcontroller having face detection and recognition capabilities. Here, the ADXL335 accelerometer sensor act as the input device which is mounted on the hand, ATmega16 microcontroller act as the processing unit, DC Motor Driver act as the driver for the motors connected to the Robot, and RF link acts as a channel for wireless communication. The ATmega16 microcontroller reads the analog output values(x and y-axis values of the accelerometer, z-axis value left, as it is not needed in this paper) of the ADXL335 accelerometer sensor and converts that analog values to digital values with its analog to digital converter. Face detection and Recognition begins with extracting the coordinates of features such as width of mouth, width of eyes, pupil, and compare the result with the measurements stored in the database and return the closest record (facial metrics). Face Recognition and Detection would be done using Principle Component Analysis Method. A camera will be mounted on the top of our main unit of the robot. The camera of our robot and the computer on which live steaming will be played will be connected to the same Wi-Fi connection. This robot will help in various application such as security of a place, attendance system etc.

Keywords-Gesture;Accelerometer;Microcontroller;Face detection;Face Recognition ;

I. INTRODUCTION

A manual robot that controls uses gesture recognition to control the movement of the robot. Gesture recognition can be defined as the manner by which gestures made by the user are recognized by the receiver module. Gestures are significant movements of the body that are associated with the physical movements of the fingers, hands, hands, face or body. In this system, the movements of the human hands are perceived by the robot through an accelerometer. As soon as the person moves the hand, the accelerometer also moves and perceives the parameters according to the position of the hand. Gestures captured by the accelerometer are sent to the comparator, which assigns the proper voltage levels to the recorded movements. The information is then transmitted to the HT12E encoder to encode the data or serialize it, and then transmit it using the RF433 MHz transmitter. In the receiver section, the RF433 MHz receiver holds the received parameters, and then the received data is decoded by the HT12D. According to the data received, we can control the robot using two DC motors in all four directions. Face detection is achieved using Viola jones face detection algorithm. The Viola Jones works on the principle of scanning a sub-window capable of detecting human faces in a input image. On the other hand, face recognition is achieved using Principle component analysis. A Face Recognition system automatically identifies the faces present in images and videos.

II. BLOCK DIAGRAM

A. *Transmitter Module*

Transmitter module consists of four components as shown in the figure. An Accelerometer is an electromechanical sensor

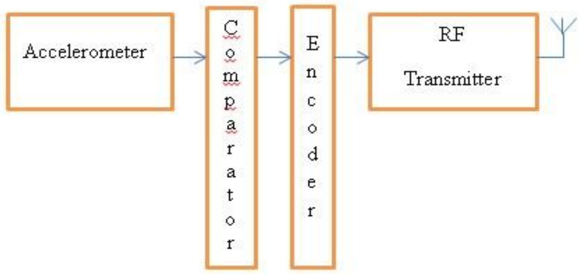

Fig.1. Block diagram of Transmitter Module

Which measures the dynamics (vibration and motion) or static (gravitational) acceleration force? It can perceive acceleration on one, two or three axes. In this system, we used the ADXL335, which is a 3-axis accelerometer. ADXl335 is a polysilicon surface microprocessor structure in which polysilicon springs suspends the structure over the surface of the plate. It also consists of a differential capacitor, which is used to measure the deviation in the structure. Any deviation in the design will unbalance the differential capacitor, leading to the output of the sensor, whose amplitude is directly proportional to the acceleration. The magnitude and direction of acceleration are measured using phase-sensitive demodulation.

The comparator is used to change the analog voltage to a digital voltage by comparing the analog voltage with a reference voltage and setting a particular high or low voltage. The radio frequency (RF) transmitter module operates at a frequency of 433 MHz and has a range of 100 m under standard conditions.

B. Receiver Module

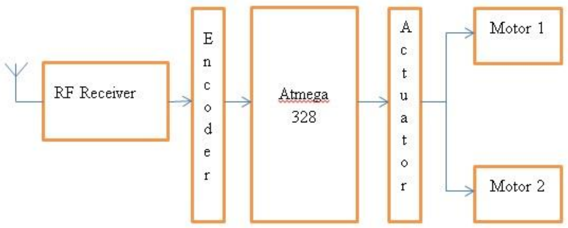

Fig. 2 Block diagram of Receiver Module

The radio frequency receiver module receives the data transmitted by the accelerometer and transmits them to the HT12D decoder. The HT12D decoder converts the received serial data into parallel data. It decodes the data only when there are no errors or inconsistent codes are found. Atmega 328 is a 28-pin high-performance Atmel 8-bit RISC microcontroller. It has 14 digital I / O pins, of which six can be used as PWM outputs and six analog input contacts. It is an integrated circuit with a processor and other support devices, such as data memory, program memory and a serial communication interface, combined together.

III. System Implimentation

A. Methodology of hand motion Recognition

A manual accelerometer is a three-dimensional solid that can be rotated around three orthogonal axes. This rotation occurs as the X axis is called a step, and the next rotation axis Y is called roll, and the last rotation around the Z axis is called yaw. Any orientation can be achieved by drawing up these three spontaneous revolutions. In our work, all the planned manual movements for robot control are simple gestures, each of which contains only one of the three rotations of the element. Gestures consisting of more than one elementary rotation are too complex for such applications.

B. Methodology of communication signal

The RF transmitter module is a small printed circuit board (PCB) capable of transmitting radio waves and modulating this wave for data transmission. Transmitter modules are usually implemented together with a microcontroller that will provide data to the module. Radio frequency transmitters are usually subject to regulatory requirements that dictate the maximum permissible transmitter output power, harmonics and band edge requirement. The RF433-RX radio frequency receiver module is a 433 MHz radio frequency receiver, receives a modulated radio signal, and then demodulates it. There are two types of radio frequency receiver module. Super-regenerative modules usually have low cost and low power, using a series of amplifiers used to extract modulated data from a carrier wave. Modules with super regeneration, as a rule, are inaccurate, since their operating frequency varies in significant volumes at temperature and supply voltage. Super heterodyne receivers, which have an advantage over super-regenerative, provide increased accuracy and stability over a wide range of voltages and temperatures. This stability comes from the fixed design of the crystal, which in turn leads to a relatively more expensive product. A radio receiver that receives the transmitted encoded data from a remote location converts the code to digital format, and the output is available to the pin number 2 of master microcontroller. It's a pin of the built-in microcontroller's art. Based on the codes given as input, master microcontroller give the command to the slave microcontroller, and the robot will move in forward direction, backward direction. It will turn left or right while moving forward or in reverse direction. And in case of bump, it will move backward and wait for the next instruction.

C. Methodology of identifying Obstacles

Ultrasonic measurement module HC-SR04 is used to detect obstacles and determine the range of obstacles from the robot. The ultrasonic sensor consists of two parts: one is a transducer creating an ultrasonic sound wave with a frequency of 40 kHz, and the other is for listening to an echo. This robot has 3 mutually perpendicular ultrasonic sensors covering three directions. HC-SR04 has 4 contacts - ground, Vcc, trigger and echo. Ground pins and Vcc sensors are connected to ground and 5 V contacts on the Ardiuno board respectively, while the trigger and echoes are connected to any digital I/O terminal on the Ardiuno board.

To create an ultrasound, you need to set Trig to high state for 10 μs. This will send an 8-stroke sound pulse that will move with the sound of the speed, and it will be received in the echo buffer.

D. Face detection system

Viola Jones face detection algorithm is implemented using four key concepts.

1. Haar like feature

Haar like features are used to detect variation in the black and light portion of the image. This computation forms a single rectangle around the detected face. Based on the color shade

near nose or forehead a contour is formed. Some commonly used Haar features are:
(i). Two rectangle feature.
(ii). Three rectangle feature.
(iii). Four rectangle feature.

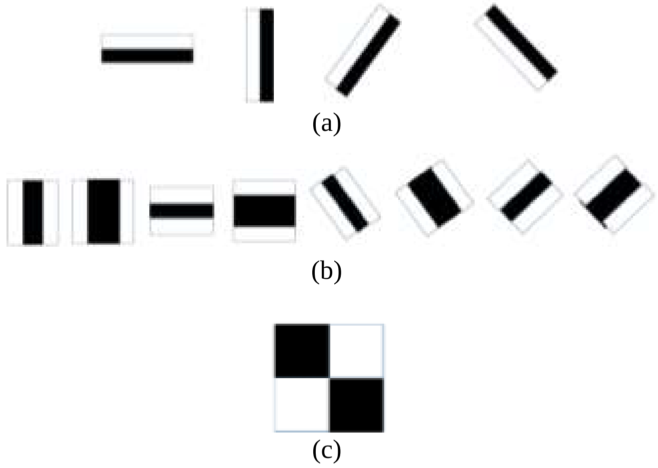

Fig. 3 (a) Two Rectangle Feature, (b) Three Rectangle Feature, (c) Four Rectangle Feature

2. Integral Image

Integral image is used to felicitate quick feature detection. The meaning of integral image is the outline of the pixel values in the original images.

1	1	1
1	1	1
1	1	1

1	2	3
2	4	6
3	6	9

Fig. 4 Integral image

3. Adaboost Algorithm

AdaBoost algorithm helps to select small features from the face that facilitates fast and easy computation .Unlike other methods, AdaBoost algorithm gives desired region of the object discarding unnecessary background. The working model can in interpreted by using neural networks.

- Given image is in the form (x1,y1)……. (xn,yn)
- y_i =0,1 for negative and positive examples.
- Initialize the weights $w_{1,i} = \frac{1}{2m}, \frac{1}{2l}$ for y_i =0,1 respectively, where m and l are number of positives and negatives respectively.
- For t=1,..,T:
  1.) Normalize the weights,
  $w_{t,i} = \frac{w_{t,i}}{\sum_{j=1}^{n} w_{t,i}}$ ....(4)
  is the probability distribution
  2.) For each feature j, train a classifier h_j which is restricted to use a single feature.The error is evaluated with respect to ,
  $w_t E_t = \sum \; w_{i|h_j(x_i) y_i|}$....(5)
  3.) Choose the classifier h_t with lowest error E_t
  4.) Update the weights
  $w_{t+1,i} = w_{t,i} B_t^{1-e_i}$....(6)
  Where e_i=0 if examples is classified correctly.
  e_i=1, otherwise
  And $B_t = \frac{e_t}{1-e_t}$
- The final strong classifier is:
  h(x) = 1 $\sum_{t=-1}^{T} a_t h_t(x) \geq \frac{1}{2} \sum_{t=1}^{T} a_t$.....(7)
  where $a_t = \log \frac{1}{Bt}$

4. Cascaded Classifiers

The Viola and Jones face detection algorithm eliminates face candidates quickly using a cascade of stages. The cascade eliminates candidates by making stricter requirements in each stage with later stages being much more difficult for a candidate to pass. Candidates exit the cascade if they pass all stages or fail any stage. A face is detected if a candidate passes all stages.

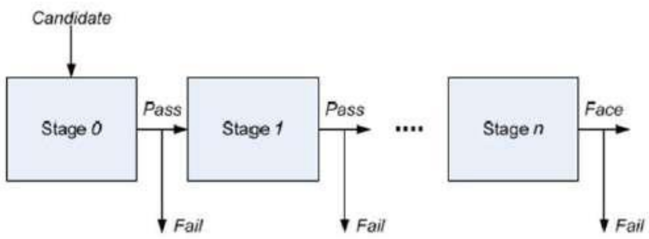

Fig. 5 Stages of Cascade Classifiers

E. Face Recognition System

The goal of a face recognition system is to discriminate input signals (image data) into several classes (persons), being important for a wide variety of problems like image and film processing, human-computer interaction, criminal identification and others.
The inputs signals can be highly noisy because of different lightning conditions, pose, expression, hair…. Nevertheless, the input signals are not completely random and even more, there are patterns present in each input signal. One can observe in all input images common objects like: eyes, mouth, nose and relative distances between these objects.
These common features are called eigenfaces in the facial recognition domain (or principal components generally). They can be extracted out of the original image data through a mathematical technique called Principal Component Analysis (PCA).

The main idea of principal component analysis is to find the vectors that best account for the distribution of face images within the entire image space. These vectors define the subspace of face images, which we call "face space". Each vector is of length N square, describes an N-by-N image, and is a linear combination of original face images, and because they are face-like in appearance, we refer then to as "eigenfaces".

Let the training set of face images be Γ1, Γ2, Γ3, Γ4…ΓM. The average face of the set is defined by

$$\Psi = 1/M \sum_{n=1}^{M} \Gamma n$$

Each face differs from the average by the vector

$$\Phi i = \Gamma i - \Psi$$

The set of very large vectors is then subject to principle component analysis, which seeks a set of M orthonormal vectors, U_n, which best describes the distribution of data. The kth vector, U_k, is chosen such that

$$\lambda = 1/M \sum_{n=1}^{M} \left(u_k^T \Phi_n\right)^2 \quad \text{.................................(1)}$$

is a maximum, subject to

$$u_l\, u_k = \delta_{lk} = \begin{cases} 1, & \text{if } l=k \\ 0, & \text{if Otherwise}\text{.......................(2)} \end{cases}$$

The vectors U_k and scalars λ_k are eigenvectors and eigenvalues, respectively, of the covariance matrix

$$C = \frac{1}{M \sum_{n=1}^{M} \Phi_n \Phi_n^T}$$

$$= AA^T$$

Where the matrix A = [Φ1 Φ2 Φ3…ΦM]. The matrix C, however, is N2 by N2, and determining the N square eigenvectors and eigenvalues is an intractable task for typical image sizes. We need a computationally feasible method to find these eigenvectors. If the number of data points in the image space is less than the dimension of the space (M<N2), there will be only M–1, rather than N2, meaningful eigenvectors. (The remaining eigenvectors will have associated eigenvalues of zero). Fortunately we can solve for the N2 dimensional eigenvectors in case by first solving for the eigenvectors of an M-by-M matrix – e.g., solving a 16 x 16 matrix rather than a 16,384 x 16,384 matrix and then taking appropriate linear combinations of the face images. Consider the eigenvectors vi of ATA such that

$$A^T A\, Vi = \mu_i v_i$$

Pre-multiplying both side by A we have

$$AA^T A\, Vi = \mu_i AVi$$

Following this analysis, we can construct the M by M matrix $L = A^T A$, where

$$\mathbf{Lmn} = \Phi_m^T \Phi_n,$$

And find the M eigenvectors, v**l**, of L.These vectors determine linear combinations of the M training set face images to from the eigenfaces U_l.

$$U_l = \sum_{k=1}^{M} v_{lk} \Phi_k$$

With this analysis the calculations are greatly reduced, from the order of the number of pixels in images (N2) to the order of the number images in the training set (M).The associated eigenvalues allow us to rank to eigenvectors according to their usefulness in characterizing the variation among the images.

IV. Components Testing

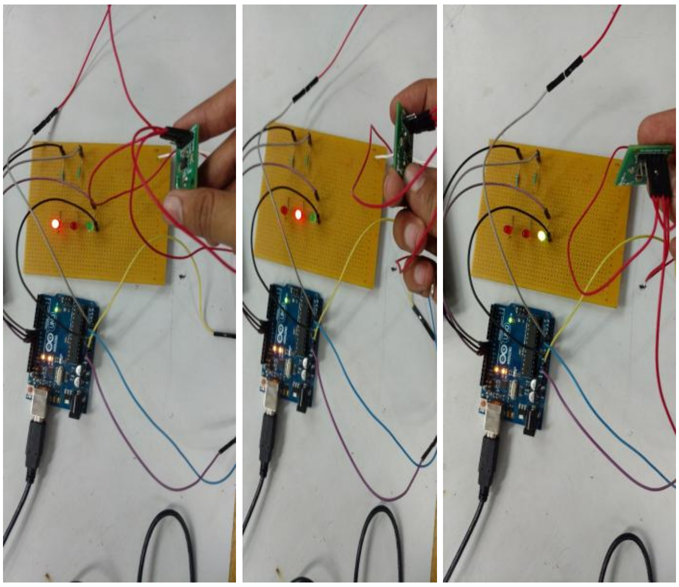

Fig. 6 Testing of Accelerometer

The Accelerometer ADXL335 was tested using an Ardiuno microcontoller and three LEDs. As shown in Fig. 3, the LED lights up according to the movement of accelerometer and ensuring proper working of ADXL335 module.

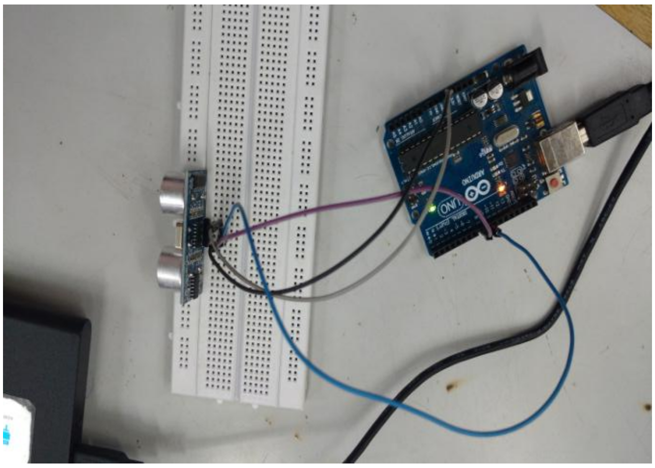

Fig. 7 Testing of Ultrasonic Sensor

V. RESULTS

In this paper, we introduced a hand-gesture-based interface for navigating a car-robot. A user can control a car-robot directly by using his or her hand trajectories. The output of three axis i.e. x-axis, y-axis and z-axis is displayed on the screen when simulated on a simulation software. When the accelerometer was tilted towards the x-axis, the value of x-coordinate changed, similarly when the hand was tilted towards y and z-axis, the values of both the axis was also changed. The change in the values of the coordinates of x-axis, y-axis and z-axis made the robot turn in desired directions. The path between the hand module and robot module was made by a transceiver circuit.

```
X=-98 Y=-302 Z=990
X=-99 Y=-300 Z=991
X=-100 Y=-302 Z=992
X=-100 Y=-302 Z=992
X=-99 Y=-304 Z=1009
X=-99 Y=-304 Z=1009
X=-99 Y=-304 Z=991
X=-99 Y=-304 Z=991
X=-99 Y=-304 Z=991
X=-102 Y=-289 Z=990
X=-98 Y=-303 Z=991
X=-98 Y=-303 Z=991
X=-98 Y=-303 Z=991
X=-99 Y=-303 Z=987
X=-100 Y=-302 Z=991
X=-98 Y=-295 Z=991
X=-98 Y=-295 Z=991
X=-99 Y=-303 Z=992
X=-98 Y=-303 Z=991
X=-98 Y=-303 Z=991
X=-98 Y=-303 Z=991
X=-108 Y=-303 Z=990
X=-99 Y=-302 Z=992
X=-99 Y=-302 Z=992
X=-99 Y=-302 Z=992
X=-99 Y=-304 Z=990
X=-99 Y=-303 Z=992
X=-94 Y=-302 Z=996
X=-94 Y=-302 Z=996
X=-94 Y=-302 Z=996
X=-100 Y=-302 Z=990
X=-100 Y=-302 Z=990
X=-100 Y=-302 Z=990
X=-97 Y=-303 Z=1005
X=-97 Y=-303 Z=991
X=-97 Y=-303 Z=991
X=-97 Y=-303 Z=991
```

Fig. 8 Output of Accelerometer

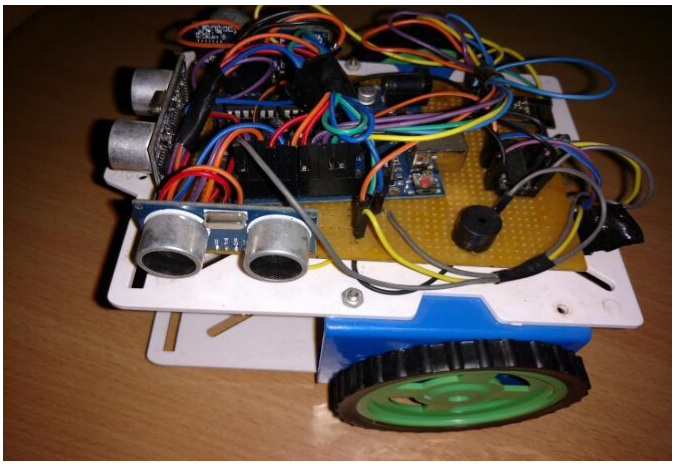

Fig. 9 Side view of Robot

Firstly, images of faces need to be inserted in the database. This is achieved by input the images of faces in the application. This is linked to the database which is used to compare and authenticate the faces it needs to recognize.

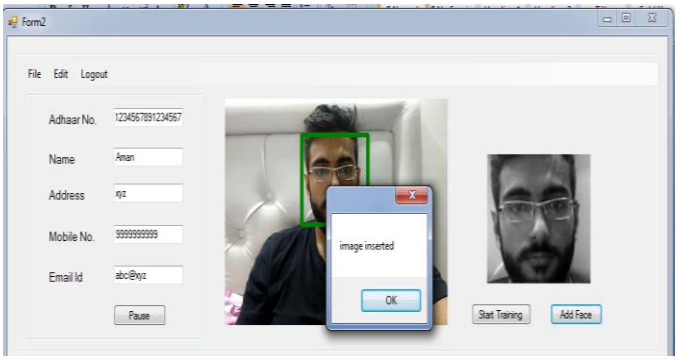

Fig. 10 Creating Database

Once the database is ready and it contains all the people it needs to recognize, we test the application for face recognition. The application first detects the face. After that, if the application compares it with the database that whether it is a known face or not .If it is a known face, then application identifies the face and gives the name of the person.

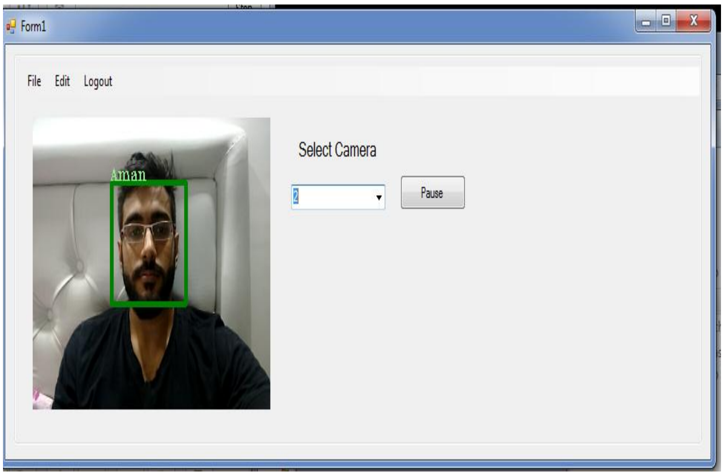

Fig. 11 Face Detection and Recognition

VI. CONCLUSION AND FUTURE SCOPE

In this paper, we have designed a hand gesture controlled robot with the capabilities of face recognition and detection. The movement of the robot car is being controlled by a 3 axis accelerometer employed on a hand with X, Y, Z values and these values are sent to the development board with robot car using RF Module. The Ultrasonic sensor is used for detecting any obstacles in the way of robot car. It reduces the physical hardship of user and provide user with an ability and freedom to maneuver the robot in a desired direction. Along with successfully using accelerometer and Ultrasonic Sensor in the robot, we have also used Viola Jones algorithm and Principle component analysis to provide face Detection and recognition capabilities, respectively. Both the approaches are robust, simple, and easy and fast to implement compared to other algorithms. They provide a practical solution to the detection and recognition problem. This technology has a vast future scope and applications such as face identification, access control, Image database investigation etc. We believe that the research in face detection and recognition is an exciting area for many years to come and will keep many scientists and engineers busy. In the future, 2D & 3D Face Recognition and large scale applications such as e-commerce, student ID, digital driver licenses, or even national ID is the challenging task in face recognition and detection & the topic is open to further research.

References

[1] A. Mulder, "Hand gestures for HCI", *Technical Report 96-1*, vol. Simon Fraster University, 1996

[2] Mayank Chauha and Mukesh Sakle, "Study & Analysis of Different Face Detection Techniques," *International Journal of Computer Science and Information Technologies*, Vol. 5 (2), 2014, 1615-1618

[3] F. Quek, "Towards a Vision Based Hand Gesture Interface", pp. 17-31, in *Proceedings of Virtual Reality Software and Technology, Singapore, 1994*.

[4] A. Lanitis, C.J. Taylor, and T.F. Cootes, "An Automatic Face Identification System Using Flexible Appearance Models," *Image and Vision Computing*, vol. 13, no. 5, pp. 393-401, 1995.

[5] A. J. Heap and D. C. Hogg, "Towards 3-D hand tracking using a deformable model", In *2nd International Face and Gesture Recognition Conference*, pages 140–145, Killington, USA, Oct. 1996.

[6] Z. Shiqi, Y. Chun and Z. Yan, "Handwritten character recognition using orientation quantization based on 3D accelerometer", the *5th Annual International Conference on Mobile and Ubiquitous Systems*, Dublin, Ireland, 25 July 2008.

[7] K. K. Kim, K. C. Kwak, and S. Y. Ch, "Gesture Analysis for Human-Robot Interaction", *Proc. of the 8th Int. Conf. on Advanced Communication Technology*, Vol. 3, pp. 1824–1827. 2006

[8] Richard Watson, "A Survey of Gesture Recognition Techniques", Technical Report TCD-CS-93-11, Department of Computer Science, Trinity College Dublin, 1993

[9] Y. Wu and T.S. Huang, "Hand modelling analysis and recognition for vision-based human computer interaction", *IEEE Signal Processing Mag. – Special issue on Immersive.Interactive Technology*, vol.18, no.3, pp. 51-60, May 2001

[10] A. J. Heap and D. C. Hogg, "Towards 3-D hand tracking using a deformable model", In 2nd International *Face and Gesture Recognition Conference*, pages 140–145, Killington, USA, Oct. 1996.

[11] K G Derpains, "A review of Vision-based Hand Gestures", Internal Report, Department of Computer Science. York University, February 2004

.