Microcontroller and Sensors Based Gesture Vocalizer

ATA-UR-REHMAN

Electrical Engineering Department
Lecturer Rachna College of Engg. and tech. Gujranwala Pakistan

SALMAN AFGHANI

Computer Engineering Department
Professor Air University Islamabad Pakistan

MUHAMMAD AKMAL

Electrical Engineering Department
Lecturer Rachna College of Engg & Tech. Gujranwala Pakistan

RAHEEL YOUSAF

Electrical Engineering Department
Air University Islamabad Pakistan

Abstract:- Gesture Vocalizer is a large scale multi-microcontroller based system being designed to facilitate the communication among the dumb, deaf and blind communities and their communication with the normal people. This system can be dynamically reconfigured to work as a "smart device". In this paper, microcontroller and sensors based gesture vocalizer is presented. Gesture vocalizer discussed is basically a data glove and a microcontroller based system. Data glove can detect almost all the movements of a hand and microcontroller based system converts some specified movements into human recognizable voice. The data glove is equipped with two types of sensors: The bend sensors and accelerometers as tilt sensors. This system is beneficial for dumb people and their hands will speak having worn the gesture vocalizer data glove.

*Key-Words:-*Gesture Recognition, Data Glove, Bend Sensor, Tilt Sensor, Speech Synthesis

1 Introduction

"Speech" and "gestures" are the expressions, which are mostly used in communication between human beings. Learning of their use begins with the first years of life. Research is in progress that aims to integrate gesture as an expression in Human-Computer Interaction (HCI).

In human communication, the use of speech and gestures is completely coordinated. Machine gesture and sign language recognition is about recognition of gestures and sign language using computers. A number of hardware techniques are used for gathering information about body positioning; typically either image-based (using cameras, moving lights etc) or device-based (using instrumented gloves, position trackers etc.), although hybrids are beginning to come about [1], [2], [3].

However, getting the data is only the first step. The second step, that of recognizing the sign or gesture once it has been captured is much more challenging, especially in a continuous stream. In fact currently, this is the focus of the research.

This research paper analyses the data from an instrumented data glove for use in recognition of some signs and gestures. A system is developed for recognizing these signs and their conversion into speech.

The results will show that despite the noise and accuracy constraints of the equipment, the reasonable accuracy rates have been achieved.

2 Previous work

Many scientists are working in the field of gesture recognition. A wonderful and latest survey of the work done in this field is described in reference [4]. Reference [5] and [6] discuss the gesture recognition for human robot interaction and human robot symbiosis. Reference [7] offers a novel "signal-level" perspective by exploring prosodic phenomena of spontaneous gesture and speech co-production. It also presents a computational framework for improving continuous gesture recognition based on two phenomena that capture voluntary (co-articulation) and involuntary (physiological) contributions of prosodic synchronization. Reference [8] discusses different categories for gesture recognition. Markov models are used for gesture recognition in reference [9] and [10]. A comprehensive framework is presented that addresses two important problems in gesture recognition systems in [11]. An augmented reality tool for vision based hand gesture recognition in a camera-projector system is described in reference [12]. A methodology using a neighborhood-search algorithm for tuning system parameters for gesture recognition is addressed in [13]. A novel method is introduced to recognize and estimate the scale of time-varying human gestures in [14].

3 Methodology

Block diagram of the system is shown Fig.1. The system is consisted of following modules:

- Data Glove
- Tilt detection
- Gesture detection
- Speech Synthesis
- LCD Display

Data glove is consisted of two sensors; bend sensors and tilt sensor. The output of the tilt sensors is detected by the tilt detection module, while the output of the bend sensors, and the overall gesture of the hand are detected by the gesture detection module. The gesture detection module gives an 8-bit address to speech synthesis module; 8-bit address is different for each gesture. Speech Synthesis module speaks the message respective to address received by it.
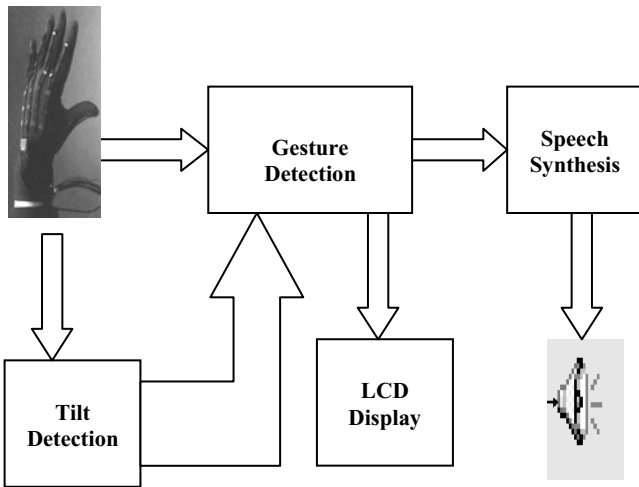

Fig.1: Block Diagram of the System

4 System Description

4.1 Data Glove

Data glove is consisted of two sensors; bend sensors and tilt sensor

4.1.1 Bend sensor

In this research setup data glove is equipped with five bend sensors, each of the bend sensor is meant to be fixed on each of the finger of the hand glove for the monitoring and sensing of static movements of the fingers of the hand.

The bend sensor is made by using 555 timer IC in astable mode along with a photo transistor. The output of the bend sensor is a square wave. Frequency of this output wave varies with the bending of the bend sensor. Circuit diagram of bend sensor is shown below in Fig.2.

Each bend sensor has its own square output which is required to be transferred to the third module of the system where pulse width of the output of each bend sensor is calculated with the help of microcontroller.
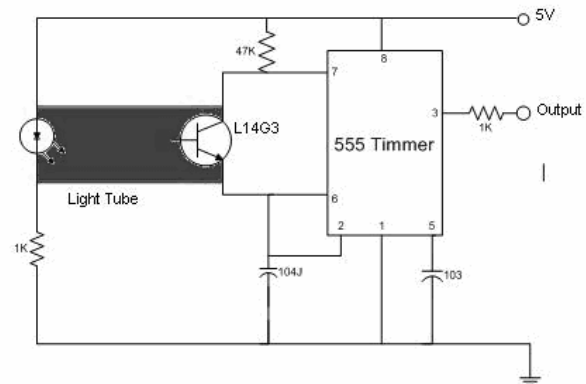

Fig. 2: Circuit Diagram of Bend Sensor

4.1.2 Tilt Sensor

Accelerometer in the Gesture Vocalizer system is used as a tilt sensor, which checks the tilting of the hand. ADXL103 accelerometer is used in the system, the accelerometer has an analog output, and this analog output varies from 1.5 volts to 3.5 volts. The output of the accelerometer is provided to third module, which includes pipeline structure of two ADC's. There is a technical issue at this stage of the project that is if we convert the analog output of the accelerometer, which ranges from 1.5 volts to 3.5 volts to a digital 8-bit output the systems become very sensitive. Reason is division of 2 volts range into 256 ($2^8 = 256$) steps is much more sensitive than converting 5 volts range into 256 steps. Now the question arises, why do we need a less sensitive system, the answer if a more sensitive system is used then there is a huge change in the digital output with the very little tilt of the hand, which is difficult to be handled.
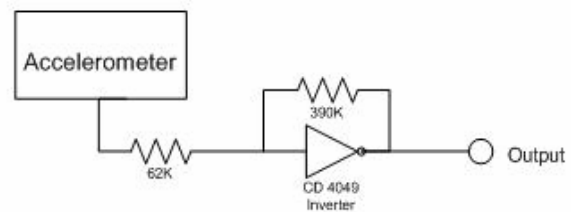

Fig. 3: Amplification and Attenuation Circuitry

Solution to this problem is to increase the range of the output of the accelerometer by using an additional circuitry. This additional circuitry increases the range of the accelerometer from 2 volts range to 5volts range.

The circuitry shown in Fig.3 is used for the amplification and attenuation purposes. Amplification means amplification of upper range i.e. 3.5 volts to the 5 volts, and the attenuation means the attenuation of lower range i.e. 1.5 volts to 0 volts.

4.2 Tilt Detection

The basic function of this module is to detect the tilting of the hand and sending some binary data against meaningful gestures, to the bend detection module.

The output, which is obtained from the accelerometers after amplification, is an analog output. To deal with this analog output, and to make it useful for the further use, it is required to change it into some form, which is detectable for the microcontroller. The analog output of the accelerometer is converted into digital form. A lot of analog to digital converter IC's are available in the market. This Gesture Vocalizer system is a dual axis system, which can detect the tilt of the hand in two axes. A dual channel ADC can be used to convert the outputs of two accelerometers in to digital form. The problem is that, low price multi-channel ADC's have a high error in their output, efficient low error ADC's are very costly.

So the ultimate solution is to use a pipeline structure of two single channel ADC's. Two ADC0804 IC's are used in this system. Both ADC's have 8-bit output.

The chip select pin of the ADC0804 is used to make a pipeline structure of the two ADC's. Both ADC's have common data bus, and common control lines. The difference is, both have separate chip select. At a time only one chip is selected. So the data bus and the control lines are used for one ADC at a time. To make sure that only one chip is selected at a time, a direct chip select signal is given to the one ADC and an inverted signal is given to the second ADC. Both ADC's are controlled through single microcontroller. The chip select signal is complemented at the end of each conversion. So at first, the first ADC converts the analog signal to the digital form and then the second ADC converts the analog signal of second accelerometer into digital form.
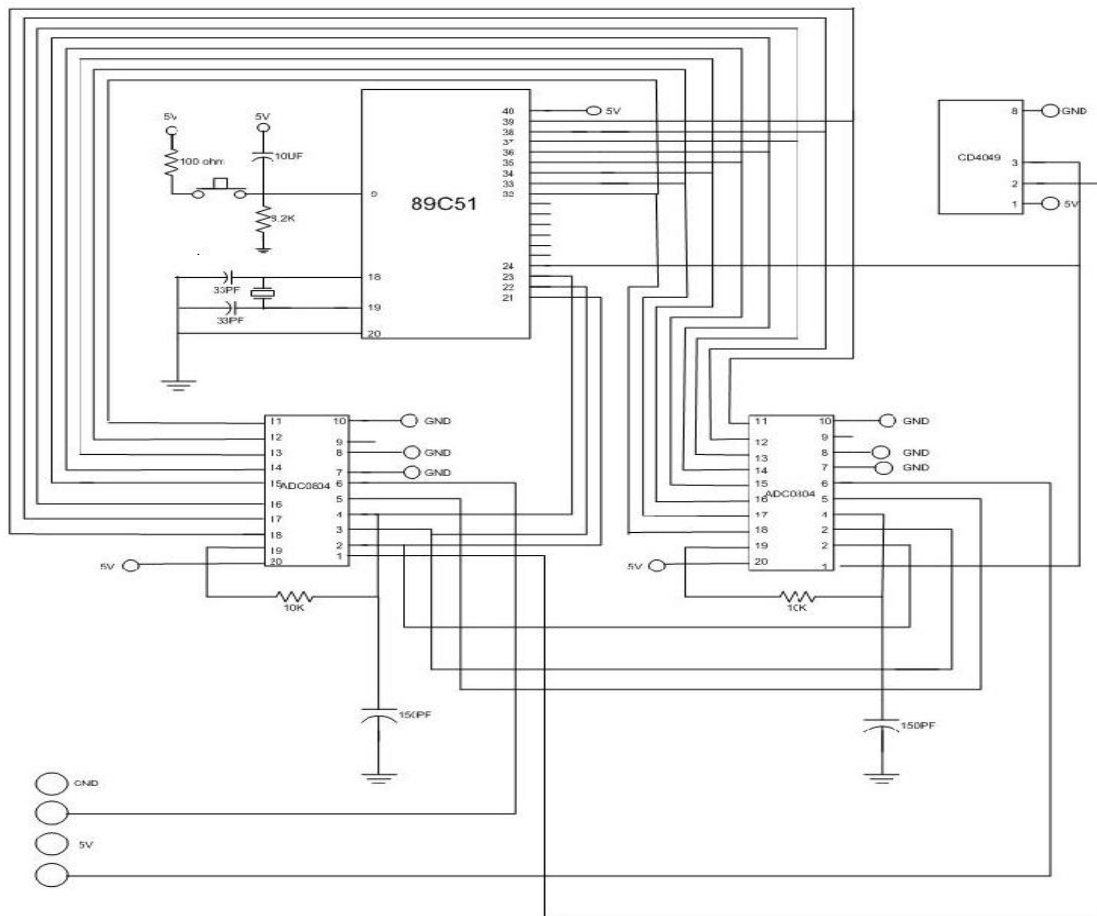

Fig.4: Pipe line structures of ADC's

Now the output of the accelerometers is converted into the digital form this output is useful, in a sense that it is detectable by the microcontroller, and useful for the further use. Fig.4 shows the complete circuit diagram of the pipeline structure of the two ADC's.

The common data bus of two ADC's is attached to a port of the microcontroller, which is controlling both the ADC's. On this data bus ADC's send converted data of their respective accelerometers to the microcontroller. Microcontroller receives the data of the two ADC's one by one, and saves them, for the further use.

Next step for the microcontroller is to check the data from the ADC's. The microcontroller checks whether the data received from the ADC's is some meaningful data, or useless one. Meaningful means that the tilt of the hand is some meaningful tilt and hand is signaling some defined gesture, or a part of the gesture, because gesture means a complete motion of the hand in which the bending of the finger is also involved. The microcontroller compares the values received from the ADC's with the predefined values, which are present in the memory of the microcontroller and on the basis of this comparison the microcontroller decides that, is the gesture a meaningful gesture.

If the hand is signaling a meaningful gesture then the microcontroller moves toward the next step. The next step of the microcontroller is to send eight bit binary data to the main "bend detection" module. The eight-bit code is different for every valid gesture. On the basis of this code, which is, sent by the tilt detection module, the "bend detection" module checks the gestures as a whole, and takes some decisions. The "bend detection module" sends eight bit data to the speech syntheses module that knows the meaning of each data.

4.3   Bend Detection

The bend detection module is the most important and the core part of the paper. This module is based on a microcontroller-controlled circuitry. In this module one microcontroller is used and three ports of this microcontroller are in use. Port zero takes the input from the five bend sensors, which is to be processed. The port one takes data from the tilt detection module and the port three gives final data, which represents some meaningful gesture to the speech synthesis module.

At first the microcontroller takes input of the five-bend sensor at its port zero. Output of the five bend sensors is given at the separate pin.

Microcontroller deals with the bend sensors one by one. First of all the microcontroller checks the output of the first bend sensor, and calculates its pulse width, after the calculation of the pulse width of the first bend sensor the microcontroller saves its output, and then moves towards the second bend sensor and calculates its pulse width in the similar manner, and keeps on calculating the pulse width of the bend sensors one by one, having calculated the pulse width of the outputs of the five bend sensors, the microcontroller moves towards the next step of the module, i.e. gesture detection.

Gesture detection is the most important part of this module. The pulse width calculation part of the module calculates the pulse width of the signal obtained from the bend sensors at a regular interval. Even a little bend of the finger is detected at this stage of the system, so the bending of the figure has infinite levels of bends, and the system is very sensitive to the bending of the finger. Now the bending of each finger is quantized into ten levels. At any stage, the finger must be at one of these levels, and it can easily be determined how much the finger is bended. So far the individual bending of each finger is captured. System knows how much each finger is bended. Now the next step is to combine the movement of each finger and name it a particular gesture of the hand.

Now the system reads the movements of five fingers as a whole, rather than reading the individual finger. Having read bending of the fingers, the system checks whether the bend is some meaningful bend, or a useless or undefined bend. If the bending of the fingers gives some meaningful gesture, then system moves towards the next step.

In the next step the system checks the data, which was sent by tilt detection module at port one of the microcontroller. The data sent by this module shows whether the tilt of the hand is giving some meaningful gesture or it is undefined. If the tilt of the hand is also meaningful then it means the gesture as a whole is a meaningful gesture.

So far it is detected by the system whether the gesture given by hand is some meaningful gesture, or a useless one. If the gesture is meaningful the system sends an eight bit data to the speech synthesis module. This eight bit data can represent 256 ($2^8=256$) different gestures. The gesture detection module assigns a different 8bit code to each gesture.

4.4   Speech Synthesis

This module of the system is consisted of a microcontroller (AT89C51), a SP0256 (speech synthesizer) IC, amplifier circuitry and a speaker.

The function of this module is to produce voice against the respective gesture. The microcontroller receives the eight bit data from the "bend detection" module. It compares the eight bit data with the predefined values. On the basis of this comparison the microcontroller comes to know that which gesture does the hand make.

Now the microcontroller knows that which data is send by the bend detection module, and what the meaning of this data is. Meaning means that the microcontroller knows, if the hand is making some defined gesture and what should the system speak.

The last step of the system is to give voice to the each defined gesture. For this purpose a speech synthesizer IC, SPO256 is used. Each word is consisted of some particular allophones and in case of Speech synthesizer IC each allophones have some particular addresses. This address is to be sent to the SPO256 at its address lines, to make the speaker, speak that particular word.

The summary of the story is that we must know the address of each word or sentence, which is to be spoken by this module. Now these addresses are already stored in the microcontroller. So far, the microcontroller knows what is the gesture made by the hand, and what should be spoken against it. The microcontroller sends the eight-bit address to SPO256. This eight-bit address is representing the allophones of the word to be spoken.

SPO256 gives a signal output. This signal is amplified by using the amplifying circuitry. The output of the amplifier is given to the speaker.

4.5 LCD Display

By using the gesture vocalizer the dumb people can communicate with the normal people and with the blind people as well, but the question arises that how can the dump people communicate with the deaf people. The solution to this problem is to translate the gestures, which are made by the hand, into some text form. The text is display on LCD. The gestures are already being detected by the "Gesture Detection" module. This module sends signal to the speech synthesis module, the same signal is sent to the LCD display module. The LCD display module is consisted of a microcontroller and an LCD. The microcontroller is controlling the LCD. A signal against each gesture is received by LCD display module. The LCD display module checks each signal, and compares it with the already stored values. On the basis of this comparison the microcontroller takes the decision what should be displayed, having taken the decision the

microcontroller send an eight bit address to the LCD, this eight bit address, tells the LCD, what should be displayed.

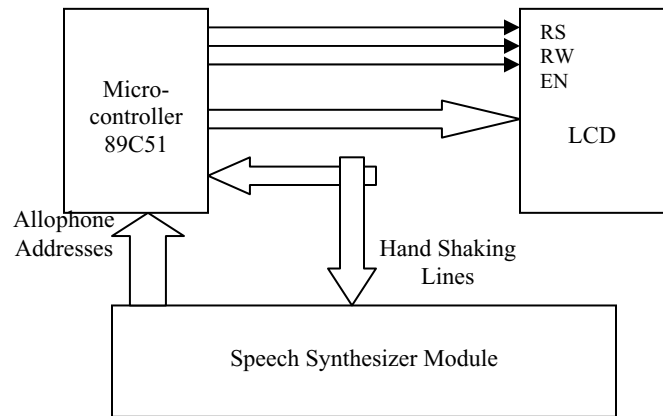The block diagram of the LCD display module is shown in the Fig.6.

Fig. 5: Block Diagram LCD Display Module

5 Over View of the System

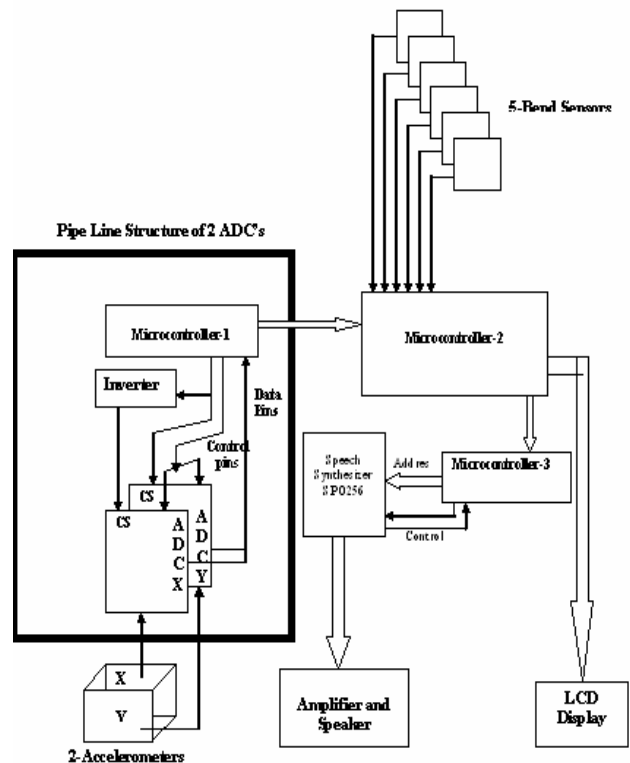The Fig.6 gives an over view of the whole project.

Fig. 6: over view of the project

X and Y in the accelerometer box shows that this system is a dual axis system and two accelerometers and two ADC's are used for each axis.

6. Conclusion and Future Enhancements

This research paper describes the design and working of a system which is useful for dumb, deaf and blind people to communicate with one another and with the normal people. The dumb people use their standard sign language which is not easily understandable by common people and blind people cannot see their gestures. This system converts the sign language into voice which is easily understandable by blind and normal people. The sign language is translated into some text form, to facilitate the deaf people as well. This text is display on LCD.

There can be a lot of future enhancements associated to this research work, which includes:

1- Designing of wireless transceiver system for "Microcontroller and Sensors Based Gesture Vocalizer".
2- Perfection in monitoring and sensing of the dynamic movements involved in "Microcontroller and Sensors Based Gesture Vocalizer".
3- Designing of a whole jacket, which would be capable of vocalizing the gestures and movements of animals.
4- Virtual reality application e.g., replacing the conventional input devices like joy sticks in videogames with the data glove.
5- The Robot control system to regulate machine activity at remote sensitive sites.

References

[1]. Kazunori Umeda Isao Furusawa and Shinya Tanaka, "Recognition of Hand Gestures Using Range Images", *Proceedings of the 1998 IEEW/RSJ International Conference on intelligent Robots and Systems Victoria, B.C.*, Canada October 1998, pp. 1727-1732

[2]. Srinivas Gutta, Jeffrey Huang, Ibrahim F. Imam, and Harry Wechsler, "Face and Hand Gesture Recognition Using Hybrid Classifiers", ISBN: 0-8186-7713-9/96, pp.164-169

[3]. Jean-Christophe Lementec and Peter Bajcsy, "Recognition of Am Gestures Using Multiple Orientation Sensors: Gesture Classification", *2004 IEEE Intelligent Transportation Systems Conference Washington, D.C.,* USA, October 34, 2004, pp.965-970

[4]. Sushmita Mitra and Tinku Acharya,"Gesture Recognition: A Survey", *IEEE TRANSACTIONS ON SYSTEMS, MAN, AND CYBERNETICS—PART C:*
APPLICATIONS AND REVIEWS, VOL. 37, NO. 3, MAY 2007, pp. 311-324

[5]. Seong-Whan Lee, "Automatic Gesture Recognition for Intelligent Human-Robot Interaction" *Proceedings of the 7th International Conference on Automatic Face and Gesture Recognition (FGR'06)* ISBN # 0-7695-2503-2/06

[6]. Md. Al-Amin Bhuiyan, "On Gesture Recognition for Human-Robot Symbiosis", *The 15th IEEE International Symposium on Robot and Human Interactive Communication (RO-MAN06)*, Hatfield, UK, September 6-8, 2006, pp.541-545

[7]. Sanshzar Kettebekov, Mohammed Yeasin and Rajeev Sharma, "Improving Continuous Gesture Recognition with Spoken Prosody*", Proceedings of the 2003 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'03),* ISBN # 1063-6919/03, pp.1-6

[8]. Masumi Ishikawa and Hiroko Matsumura, "Recognition of a Hand-Gesture Based on Self-organization Using a Data Glove", ISBN # 0-7803-5871-6/99, pp. 739-745

[9]. Byung-Woo Min, Ho-Sub Yoon, Jung Soh, Yun-Mo Yangc, and Toskiaki Ejima, "Hand Gesture Recognition Using Hidden Markov Models", ISBN # 0-7803-4053-1/97, pp.4232-4235

[10]. Andrew D. Wilson and Aaron F. Bobick, "Parametric Hidden Markov Models for Gesture Recognition", *IEEE TRANSACTIONS ON PATTERN ANALYSIS AND MACHINE INTELLIGENCE,* VOL. 21, NO. 9, SEPTEMBER 1999, pp. 884-900

[11]. Toshiyuki Kirishima, Kosuke Sato and Kunihiro Chihara, "Real-Time Gesture Recognition by Learning and Selective Control of Visual Interest Points", *IEEE TRANSACTIONS ON PATTERN ANALYSIS AND MACHINE INTELLIGENCE,* VOL. 27, NO. 3, MARCH 2005, pp. 351-364

[12]. Attila Licsár and Tamás Szirány, "Dynamic Training of Hand Gesture Recognition System", *Proceedings of the 17th International Conference on Pattern Recognition (ICPR'04),* ISBN # 1051-4651/04,

[13]. Juan P. Wachs, Helman Stern and Yael Edan, "Cluster Labeling and Parameter Estimation for the Automated Setup of a Hand-Gesture Recognition System", *IEEE TRANSACTIONS ON SYSTEMS, MAN, AND CYBERNETICS—PART A: SYSTEMS AND HUMANS,* VOL. 35, NO. 6, NOVEMBER 2005, pp. 932-944

[14]. Hong Li and Michael Greenspan, "Multi-scale Gesture Recognition from Time-Varying Contours", *Proceedings of the Tenth IEEE International Conference on Computer Vision (ICCV'05)*