IOSR Journal of Electronics and Communication Engineering (IOSR-JECE)
ISSN: 2278-2834, ISBN: 2278-8735, PP: 19-26
www.iosrjournals.org

Gesture Recognition-A Review

[1]Miss. Kawade Sonam P, [2] Prof. V. S. Ubale

[1]*M.E. Student, Electronics Dept., AVCOE, Sangamner.*
[2]*Assistant Professor, Electronics Dept., AVCOE, Sangamner.*

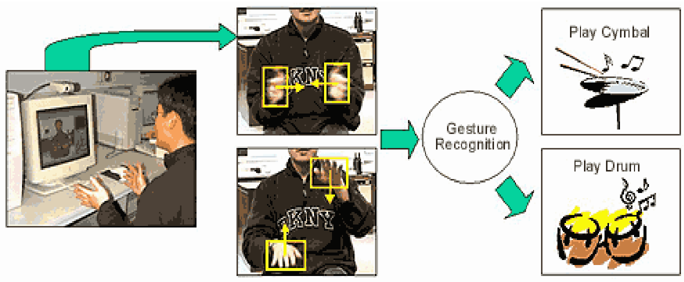
ABSTRACT- *Gesture Recognition means identification and recognition of gestures originates from any type of body motion but commonly originate from face or hand. Current focuses in the field include emotion recognition from the face and hand gesture recognition .Gesture recognition enables humans to interface with the machine* ***(HMI)*** *and interact naturally without any mechanical devices. The applications of gesture recognition are manifold, ranging from sign language through medical rehabilitation to virtual reality. This paper focuses on the Gesture recognition concept, gesture types and different ways of gesture recognition.*
Keywords - *Image Professing and Pattern Recognition.*

1. INTRODUCTION

1.1 Gesture Recognition Technology

In gesture recognition technology, a camera reads the movements of the human body and communicates the data to a computer that uses the gestures as input to control devices or applications.

For example, a person clapping his hands together in front of a camera can produce the sound of cymbals being crashed together when the gesture is fed through a computer.



Photograph 1.1: Gesture Recognition Technology

One way gesture recognition is being used is to help the physically impaired to interact with computers, such as interpreting sign language. The technology also has the potential to change the way users interact with computers by eliminating input devices such as joysticks, mice and keyboards and allowing the unencumbered body to give signals to the computer through gestures such as finger pointing.

Unlike haptic interfaces, gesture recognition does not require the user to wear any special equipment or attach any devices to the body. The gestures of the body are read by a camera instead of sensors attached to a device such as a data glove.

In addition to hand and body movement, gesture recognition technology also can be used to read facial and speech expressions (i.e., lip reading), and eye movements. The review includes ongoing work in the computer vision field on capturing gestures or more general human pose and movements by cameras connected to a computer.

II. GESTURE RECOGNITION

2.1Gesture recognition :

Gesture recognition is the process by which gestures made by the user are used to convey the information or for device control. In everyday life, physical gestures are a powerful means of communication. A set of physical gestures may constitute an entire language, as in sign languages. They can economically convey a rich set of facts and feelings.

This seminar makes the modest suggestion that gesture-based input is such a beneficial technique to convey the information or for device control with the help of identification of specific human gestures.

Research into the uses of gesture in human computer interaction is embryonic, and we hope to have inspired others to exercise their ingenuity in developing effective gestures.

A primary goal of Gesture recognition research is to create a system which can identify specific human gestures and use them to convey information or for device control. Interface with computers using gestures of the human body, typically hand movements. In gesture recognition technology, a camera reads the movements of the human body and communicates the data to a computer that uses the gestures as input to control devices or applications.

Gesture recognition is a topic in computer science and language technology with the goal of interpreting human gestures via mathematical algorithms. Gestures can originate from any bodily motion or state but commonly originate from the face or hand. Current focuses in the field include emotion recognition from the face and hand gesture recognition. Many approaches have been made using cameras and computer vision algorithms to interpret sign language. However, the identification and recognition of posture, gait, proxemics, and human behaviors is also the subject of gesture recognition techniques. The main objective of Gesture Recognition Review is:

→ Study of various types of Gestures
→ Study different ways of Gesture Recognition
→ Analysis of different ways of Gesture Recognition
→ Applications of Gesture Recognition

Gesture Recognition is the act of interpreting motions to determine such intent. The specific human gestures can identify using the gesture recognition technology and used to convey the various information or for various applications by controlling devices.

There are different types of gestures such as hand, face (emotion), body gestures etc. To identify and recognize these gestures there are different ways of gesture recognition such as:

→ Hand Gesture Recognition
→ Face (Emotion) Gesture Recognition
→ Body Gesture Recognition

2.2 Block Diagram:

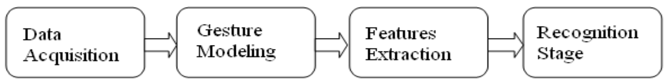


Figure 2.2: Block diagram of Gesture Recognition System

Gesture recognition system composed of several stages; these stages are varied according to application used, but, however, the unified outline can be settled, Figure 2.2 above fulfils this requirement.
A. Data Acquisition: This step is responsible for collecting the input data which are the hand, Face or Body gestures and classifier classifies the input tested gesture into required one of classes.

B. Gesture Modeling: This employed the fitting and fusing the input gesture into the model used; this step may require some pre-processing steps to ensure the successful unification.

C. Feature Extraction: After successful modeling of input data/gesture, the feature extraction should be smooth since the fitting is considered the most difficult obstacles that may face; these features can be hand/palm/fingertips location, joint angles, or any emotional expression or body movement. The extracted features might be stored in the system at training stage as templates or may be fused with some recognition devices such as neural network, HMM, or decision trees which have some limited memory should not be overtaken to remember the training data.

D. Recognition Stage: This stage is considered to be a final stage for gesture system and the command/meaning of the gesture should be declared and carried out, this stage usually has a classifier that can attach each input testing gesture matching class.

2.3 Ways of Gesture Recognition:

There are different ways of gesture recognition such as Hand, Face and Body Gesture Recognition explained below:

1) Hand Gesture Recognition
   a) Glove-based Hand Gesture Recognition

A glove-based system requires the user to be connected to the computer. Even wireless systems currently offered in the market require the user to wear a glove. Furthermore, accurate devices are expensive and hard to calibrate.

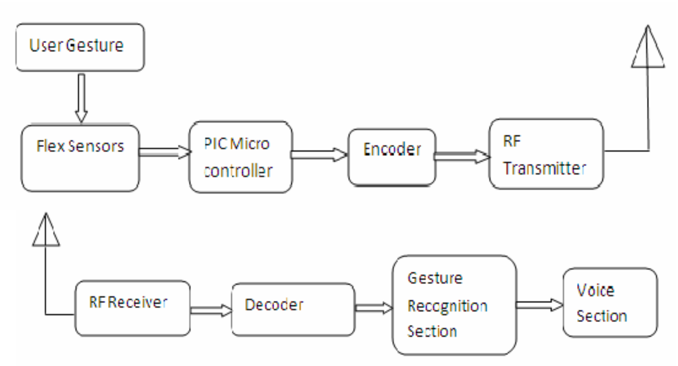


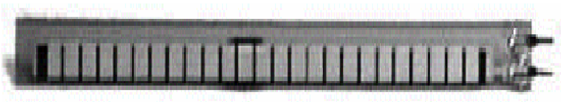
Figure 2.3.1: Block diagram of Glove-based Hand Gesture Recognition System

- → User Gesture: Data glove is implemented to capture the hand gestures of a user.
- → Flex Sensors: The data glove is fitted with flex sensors along the length of each finger and the thumb. The flex sensors output a stream of data that varies with degree of bend. The analog outputs from the sensors are then fed to the PIC microcontroller.
- → PIC microcontroller: It processes the signals and perform analog to digital signal conversion.
- → Encoder and RF Transmitter: The resulting digital signal is encoded and transmitted through RF system.
- → RF Receiver and Decoder: RF receivers receive the signal and fed to the gesture recognition section through the decoder.
- → Gesture Recognition Section: In this section the gesture is recognized and the corresponding text information is identified.
- → Voice Section: Text to speech conversion takes place in the voice section and play out through the speaker.

2.3.1 Flex Sensors:

Flex sensors are normally attached to the glove using needle and thread. They require a 5-volt input and output between 0 and 5 V, the resistivity varying with the sensor's degree of bend and the voltage output changing accordingly. The sensors connect to the device via three pin connectors (ground, live, and output). The device can activate the sensors from sleep mode, enabling them to power down when not in use and greatly decreasing power consumption.

The flex sensor Photograph: 3.1 changes resistance when bent. It will only change resistance in one direction. An unflexed sensor has a resistance of about 10,000 ohms. As the flex sensor is bent, the resistance increases to 30- 40 kilo ohms at 90 degrees. The sensor measures ¼ inch wide, 4-1/2 inches long and 0.19 inches thick.



Photograph 2.3.1: Flex Sensor

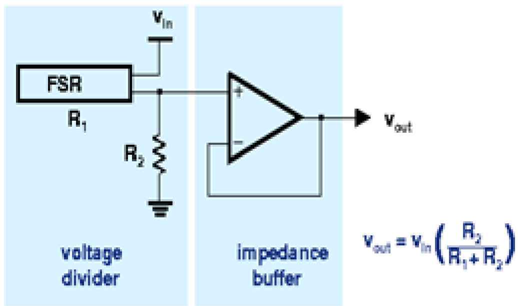
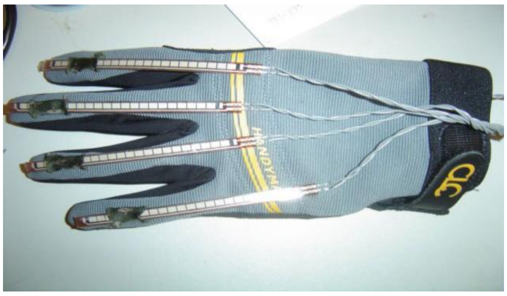


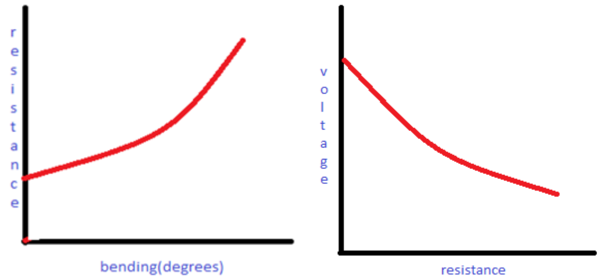
Figure 2.3.2: Basic Flex Sensor Circuit

In this two or three sensors are connected serially and the output from the sensors is inputted to the analog to digital converter in the controller. The outputs from the flex sensors are inputted into LM258/LM358 op-amps and used a non-inverted style setup to amplify their voltage. The greater the degree of bending, the lower the output voltage. The output voltage is determined based on the equation Vin * R1 / (R1 + R2), where R1 is the other input resistor to the non-inverting terminal. Using the voltage divider concept the output voltage is determined and it ranges from 1.35V to 2.5V.



Photograph 2.3.2: Glove with Flex Sensors

Characteristics:



a) b)

Figure 2.3.3: a) Bending Vs Resistance and b) Resistance Vs Voltage

PIC Microcontroller:

All output signals generated from flex sensors are in analogue form and these signals need to be digitized before they can be transmitted to encoder. Therefore microcontroller PIC16F877A is used as the main controller. It has inbuilt ADC module, which digitizes all analogue signals from the sensors and inbuilt multiplexer for sensor signal selection. It supports both serial and parallel communication facilities.

Encoder/Decoder:

The output from the PIC microcontroller is encoded by using HT12E-212 series of encoder. It is used to correct the error at the receiver end, if any error had occurred. In the receiver it is decoded by using HTI2D212 series of decoder.

Gesture Recognition Section:

Sensor data are recognized and then recorded while a user performs various sign, correlating these with specific signs and mapping them to a database. The system stores sensor data in an array for recognition. When the sensor data matches the set of values associated with a sign system recognizes that sign and output it as text. Here the microcontroller used is AT89S51.

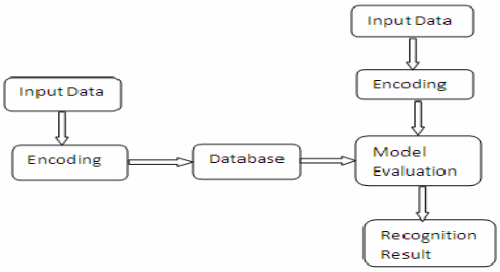


Figure 2.3.4: Model of Gesture Recognition Section

Voice Section:

Once the sensor data is matched with the database then the result of that particular sign will appear as output in the text form. This text output is given to the voice section. The speech of each text is pre-recorded and will only play out through speaker if the sign is matched.

b) Vision-based Hand Gesture Recognition

It uses one or more camera to record images of human hand gestures and lighting conditions that enhance gesture classification accuracy. There are different types of Vision-based Hand Gesture Recognition such as:

i. Infrared-camera Based
ii. Mono-camera Based
iii. Multi-camera Based Gesture Recognition

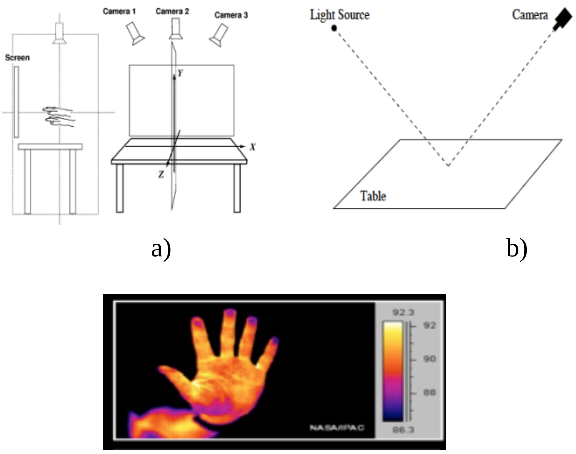


Figure 2.3.5: a) Multi-camera, b) Mono-camera and c) Infrared-camera based Gesture Recognition

Hand gestures are a means of communication, similar to spoken language. The production and perception of gestures can thus be described using a model commonly found in the field of spoken language recognition. An interpretation of this model, applied to gestures, is depicted in Figure 2.3.6.

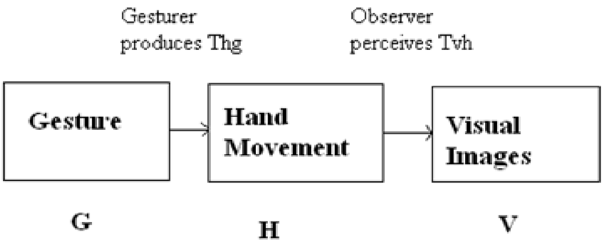


Figure 2.3.6: Production and perception of gestures

Hand gestures originate as a mental concept *G*, are expressed (*Thg*) through arm and hand motion *H*, and are perceived (*Tvh*) as visual images *V*.

According to the model, gestures originate as a gesturer's mental concept, possibly in conjunction with speech. They are expressed through the motion of arms and hands, the same way speech is produced by air stream modulation through the human vocal tract. Also, observers perceive gestures as streams of visual images of gesturers which are acquired by one or more video cameras and interpret using the knowledge they possess about those gestures. The production and perception model of gestures can also be summarized in the following form:

$$H = ThgG \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ (1)$$
$$V = TvhH \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots.\ (2)$$
$$V = Tvh(ThgG) = TvgG\ldots\ldots\ldots.\ (3)$$

Transformations *T*. can be viewed as different models: *Thg* is a model of hand or arm motion given gesture *G*, *Tvh* is a model of visual images given hand or arm motion *H*, and *Tvg* describes how visual images *V* are formed given some gesture *G*. The aim of visual interpretation of hand gestures is to infer gestures *G* from their visual images *V* using a suitable gesture model *Tvg*, or

$$\hat{G}=T^{-}vg\ V\ldots\ldots\ldots\ldots\ldots\ (4)$$

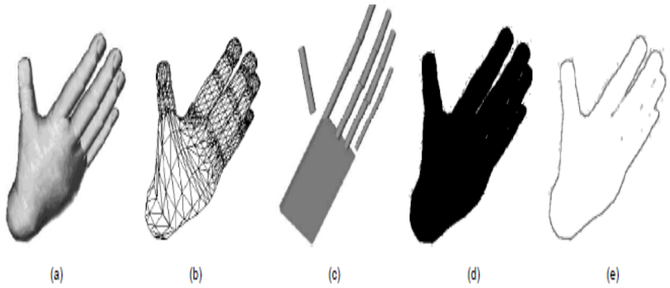
Gesture Modeling Approaches:

→ Appearance-based model: If the gesture production and perception model is considered, two possible approaches to gesture modeling may become obvious. One approach may be to try to infer gestures directly from the visual images observed. This approach has been often used to model gestures, and is usually denoted as appearance- based modeling.

→ 3D-based Model: Another approach may result if the intermediate tool for gesture production is considered: the human hand and arm. In this case, a two step modeling process may be followed:

$$\hat{H}=T^{-}vhV\ldots\ldots\ldots\ldots\ldots\ldots\ (5)$$
$$\hat{G}=T^{-}hgH\ldots\ldots\ldots\ldots\ldots\ldots\ (6)$$

In other words, one can first model the motion and posture of the hand and arm $\hat{H}$ and then infer gesture $\hat{G}$ from the motion and posture model parameters. A group of models which follows this approach is known as 3D-model-based.



Photograph 2.3.3: Hand models: (a) 3D Textured volumetric model. (b) 3D wireframe volumetric model. (c) 3D skeletal model. (d) Binary silhouette. (e) Contour

2) Face Gesture Recognition

This way of Gesture Recognition consists of:

i. OpenCV:

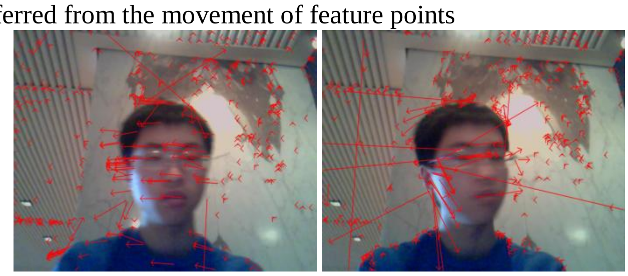
OpenCV (Free Open Source Computer Vision) is a programming functions library aimed at developing applications based on real time computer vision technologies. Some example applications in OpenCV library are Object Identification, Segmentation and Recognition, Face Recognition, Gesture Recognition, Motion Tracking, Mobile Robotics etc.

ii. Optical Flow:

Optical flow uses the pixel movement of two consecutive images. This method takes two concoctive frames and specifies how much each image pixel has been moved.

→ Estimate apparent motion in two consecutive frames
  - Find feature points in frame
  - Track the features from one frame to the next

→ Motion can be inferred from the movement of feature points



Photograph 2.3.4: Optical Flow Method

iii. Support Vector Machines:

The basic Support Vector Machine (SVM) takes a set of input data and predicts, for each given input, which of two possible classes forms the output, making it a non-probabilistic binary linear classifier.

A support vector machine constructs hyperplane or set of hyperplanes in a high or infinite-dimensional space, which can be used for classification, regression, or other tasks. Intuitively, a good separation is achieved by the hyperplane that has the largest distance to the nearest training data point of any class (so-called functional margin), since in general the larger the margin the lower the generalization error of the classifier.

Linear SVM: Given some training data $\mathcal{D}$, a set of *n* points of the form

$$\mathcal{D} = \left\{ (\mathbf{x}_i, y_i) \mid \mathbf{x}_i \in \mathbb{R}^p,\ y_i \in \{-1, 1\} \right\}_{i=1}^{n}$$

Where, y_i is either 1 or −1, indicating the class to which the point $\mathbf{X_i}$ belongs. Each X_i is a *p*-dimensional real vector. We want to find the maximum-margin hyperplane that divides the points having y_i =1 from those having y_i =-1. Any hyperplane can be written as the set of points X satisfying-

$$\mathbf{w} \cdot \mathbf{x} - b = 0,$$

where, w-normal vector and $\frac{b}{\|\mathbf{w}\|}$ determines the offset of the hyperplane from the origin along the normal vector. If the training data are linearly separable, we can select two hyperplanes, the region bounded by them is called “Margin“. Equations used:

$$\mathbf{w} \cdot \mathbf{x} - b = 1 \text{ and}$$
$$\mathbf{w} \cdot \mathbf{x} - b = -1.$$

By using geometry, we find the distance between these two hyperplanes is $\frac{2}{\|\mathbf{w}\|}$, so we want to minimize $\|\mathbf{w}\|$. As we also have to prevent data points from falling into the margin, we add the following constraint: for each i either

$$\mathbf{w} \cdot \mathbf{x}_i - b \geq 1 \qquad \text{for } \mathbf{x}_i \text{ of the first class}$$

Or $$\mathbf{w} \cdot \mathbf{x}_i - b \leq -1 \qquad \text{for } \mathbf{x}_i \text{ of the second class}$$

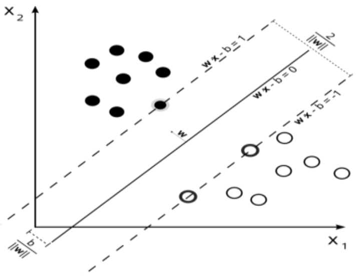


Figure 2.3.7: Maximum-margin hyperplane and margins for an SVM trained with samples from two classes.

Samples on the margin- support vectors.

3) Body Gesture Recognition:

Yang developed a system capable to detect body He does so by using a **Hidden Markov Model**.

- Body Feature Detection
- Pose Reconstruction Method: To detect human subjects out of video frames and creates a 3D representation of it. To do so 2D frames of a human being are taken by different angles.
- Least square minimization method: 2D shapes of the human are then matched with a 3D Model of human body. It allows to allocate each upper shoulder, elbow, wrist, knee, ankle etc. to their position in 3D space
- Body Feature Extraction and Gesture recognition: The previously presented 3D model gains for each frame of a video the structural feature points of the body (wrist, el- bow, shoulder, etc.). The center of this 3D space is located in the region of the trunk of the 3D human model. Each of these points are first projected into the x, y and z plane. Then the angle between this projection points and the axis are measured. These angles give the feature-

$$F_k = (\Theta x; \Theta y; \Theta z)$$

For each of this feature, feature vector can build-

$$X_t = [F_{L\text{-shoulder}}, F_{L\text{-elbow}}, F_{L\text{-wrist}} \ldots]$$

- Hidden Markov Model: Used to extract touching a knee and wrist, rising a right hand, walking, waving a hand, running, sitting on the floor, lying down on the floor, jumping and getting down on the floor gestures. It consists of two parts:
- Ergonic model: Has the task to extract the garbage movement. In this category all movements which are not to be detected are put in it.
- Left-right model: Detects the desired gestures (see Figure 2.3.8). By training the model with help of the feature vectors, gesture can be extracted.
- Reliability can be calculated as-
  Reliability = (correctly recognized gesture) ÷ (deletion error + insertion error), of over 89% for each of the movement.

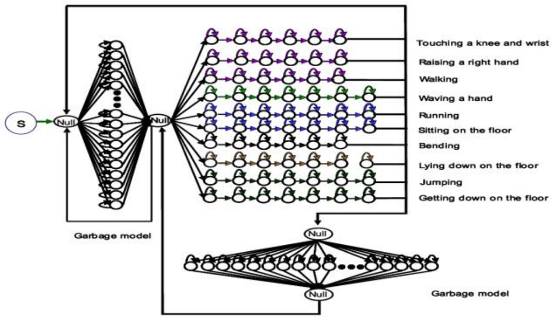


Figure 2.3.8: Key Gesture Spotting Model

III. CONCLUSIONS

3.1Conclusion

1) 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 2) interacting with the environment.
2) Gesture recognition is an extensively developed technology available designed to identify human position, action, and manipulation. Gestures are used more and more to facilitate communication with digital applications because of their expressive nature.
3) Among the various ways of gesture recognition like Hand, Face and Body Gesture Recognition, Hand Gesture Recognition is efficient technique to recognize human gestures due to its simple and greater accuracy features.

3.2. Future Scope

1) Gestures are destined to play an increasingly important role in human-computer interaction in the future.
2) Facial Gesture Recognition Method could be used in vehicles to alert drivers who are about to fall asleep.
3) Area of Hand gesture based computer human interaction is very vast. Hand recognition system can be useful in many fields like robotics, computer human interaction and so make hand gesture recognition offline system for real time will be future work to do.
4) Support Vector Machine can be modified for reduction of complexity. Reduced complexity provides us less computation time so we can make system to work real time.

3.3 Applications

1) Sign Language
2) Desktop and Tablet PC Applications
3) Robotics
4) Games
5) Directional indication through pointing
6) Monitoring automobile driver's alertness/drowsiness levels etc.

REFERENCES

[1] 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, Page No. 311-317.
[2] Vladimir I. Pavlovic, Rajeev Sharma and Thomas S. Huang, "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, Page No. 677-682.
[3] Ms. Sweta A. Raut and Prof. Nitin J. Janwe, "A Review of Gesture Recognition Using Image Object Comparison and Neural Network", International Journal of Computer Technology and Electronics Engineering (IJCTEE) National Conference on Emerging Trends in Computer Science and Information Technology (NCETSIT-2011), Page No.57-58.
[4] Mokhtar M. Hasan and Pramod K. Mishra, "Hand Gesture Modeling and Recognition using Geometric Features: A Review", Canadian Journal on Image Processing and Computer Vision Vol. 3 No. 1, March 2012, Page No. 13-16.
[5] Shoaib Ahmed.V, "MAGIC GLOVES (Hand Gesture Recognition and Voice Conversion System for Differentially Able Dumb People)".
[6] Cédric Graf, "Arm and Body gesture recognition"
[7] Hatice Gunes and Massimo Piccardi, "Automatic Temporal Segment Detection and Affect Recognition From Face and Body Display", IEEE TRANSACTIONS ON SYSTEMS, MAN, AND CYBERNETICS—PART B: CYBERNETICS, VOL. 39, NO. 1, FEBRUARY 2009, Page No. 70-72.