Gesture Recognition Technology: A Review

PALLAVI HALARNKAR
Department of Computer Engineering, NMIMS University, Vile Parle (West)
Mumbai, Maharashtra 400056, India
pallavi.halarnkar@nmims.edu

SAHIL SHAH
Department of Computer Engineering, NMIMS University, Vile Parle (West)
Mumbai, Maharashtra 400056, India
sahil0591@gmail.com

HARSH SHAH
Department of Computer Engineering, NMIMS University, Vile Parle (West)
Mumbai, Maharashtra 400056, India
harsh1506@hotmail.com

HARDIK SHAH
Department of Computer Engineering, NMIMS University, Vile Parle (West)
Mumbai, Maharashtra 400056, India
hardikshah2711@gmail.com

JAY SHAH
Department of Computer Engineering, NMIMS University, Vile Parle (West)
Mumbai, Maharashtra 400056, India
jay.shah309@gmail.com

Abstract:

Gesture Recognition Technology has evolved greatly over the years. The past has seen the contemporary Human – Computer Interface techniques and their drawbacks, which limit the speed and naturalness of the human brain and body. As a result gesture recognition technology has developed since the early 1900s with a view to achieving ease and lessening the dependence on devices like keyboards, mice and touchscreens. Attempts have been made to combine natural gestures to operate with the technology around us to enable us to make optimum use of our body gestures making our work faster and more human friendly. The present has seen huge development in this field ranging from devices like virtual keyboards, video game controllers to advanced security systems which work on face, hand and body recognition techniques. The goal is to make full use of the movements of the body and every angle made by the parts of the body in order to supplement technology to become human friendly and understand natural human behavior and gestures. The future of this technology is very bright with prototypes of amazing devices in research and development to make the world equipped with digital information at hand whenever and wherever required.

Keywords: Human-Computer Interaction; Gesture Modeling; Gesture Analysis; Gesture Recognition; Since The Beginning; The Present; The Future; Sixth Sense; 3D hand arm models; Appearance based models; Hidden Markov Models.

1. Introduction

We come across Gesture Recognition in Human–Computer Interaction. It involves the use of natural hand gestures to control devices. A gesture recognition system comprehends human gestures with mathematical algorithms. With gesture recognition, computers can become familiarized with the way humans communicate using gestures. Thus, machines and humans can interact freely with each other. The primary goal of gesture recognition is to create a system which understands human gestures uses them to control various devices. It is necessary to yield a robust and reliable system [1].

With the development of ubiquitous computing, current user interaction approaches with keyboard, mouse and pen are not sufficient. Due to the limitation of these devices the useable command set is also limited. Direct use of hands as an input device is an attractive method for providing natural Human Computer Interaction which has evolved from text-based interfaces through 2D graphical-based interfaces, multimedia-supported interfaces, to fully-fledged multi-participant Virtual Environment (VE) systems. Imagine the human-computer interaction of the future: A 3D application where you can move and rotate objects simply by moving and rotating your hand – all without touching any input device [2].

1.1 *Since the beginning*

1.1.1 *Data glove*

The first commercially available hand tracker device was the Data Glove. It uses thin fibre optic cables running down the back of each hand, each with a small crack in it. Light is shone down the cable so when the fingers are bent light leaks out through the cracks. Measuring light loss gives as accurate reading of hand pose [3].

1.1.2 *Cyber Glove*

Cyber glove developed by Kramer in 1989 uses strain gauges placed between the fingers to measure abduction as well as more accurate bend sensing. It can detect sideways movement of the fingers. But it is more expensive then the data glove [3].

1.1.3 *Theramin*

Theramin, which was developed in 1920's, is an electrical musical instrument, which responds to hand motions using two proximity sensors, one vertical, and the other horizontal. Proximity to the vertical sensor controls the music pitch, to the horizontal one loudness [3].

1.1.4 *Videoplace*

Videoplace was developed by Myron Krueger in the late 1970's. It recognizes the dynamic natural gestures, meaning users require no training. It uses real time image processing of live video of the user. It has a very good feature recognition technique which can easily distinguish between hands and fingers, whether fingers are extended or closed and even which fingers [3].

1.1.5 *Media Room*

It was the first interface to support combined speech and gesture recognition technique. It was developed by Richard Bolt in 1984. Within the Media Room the user could use speech, gesture, eye movements or a combination of all three to add, delete and move graphical objects shown on the wall projection panel. The computer interpreted the user's intentions by speech and gesture recognition and by taking the current graphical situation into account [3].

1.2 *The Present*

1.2.1 *Virtual keyboards*

Virtual keyboards use a lens to project an image of a keyboard on to a desk or other flat surface. Users then type in the virtual keyboards. An infrared light beam that the device directs above the projected keyboard detects the users fingers. The device monitor calculates how long it takes a pulse of infrared light to reflect off the users moving fingertips in return to a sensor [8].

1.2.2 *Navigaze*

Users can work with applications by moving cursors with head movements and clicking the mouse with eye blinks. Disabled people can use only the eyes for choosing an icon or file. Navigaze can recognize the difference between open closed eyes and thus respond eye blinks [8].

1.3 *The Future*

1.3.1 *CePal*

Cepal is a developed by Dhirubhai Ambani Institute of Information and Communication Technology, Gandhinagar, India. Cepal is a gizmo which can be worn like a watch and an infrared gesture based remote control which helps the motor impaired and other limb related disorders to complete routine tasks such as operating TV, air conditioner, lights and fans. It helps the people with cerebral palsy to be self-reliant [4].

1.3.2 *ADITI*

ADITI was developed by IIT-Madras. It helps people with debilitating diseases – such cerebral palsy and severe muscular skeletal disorders – to communicate using simple gestures. ADITI is an indigenous USB device that senses movement within a six-inch radius. ADITI is a screen-based device running software that provides them with a list of visual options like words pictures or alphabets to express themselves. Using ADITI patients can communicate through simple gestures like nod of heads moving feet to generate a mouse click [4].

2. Gesture Recognition System

Visual interpretation of hand gestures is mainly divided into three parts:

1) Gesture modeling
2) Gesture analysis
3) Gesture recognition

2.1 *Gesture Modeling*

It is important to first consider what models have been used for the hand gesture. The scope of a gestural interface for Human-Computer Interaction is directly related to the proper modeling of hand gestures [5].

2.1.1 *Temporal Modeling of gestures*

Three phases make a gesture [5]:

- Preparation - Sets the hand in motion from resting position.
- Nucleus - Some definite form and enhanced dynamic qualities.
- Retraction - Hand returns to the resting position or repositions for new gesture phase.

2.1.2 *Spatial Modeling of gestures*

3D hand model based models of gestures use articulated models the human hand and arm to estimate the hand and arm parameters. Appearance based models directly link the appearance of the hand and arm movements in visual images to specific gestures [5].

2.1.2.1 *3D Hand Arm Models*

These models are the premier choice of hand gesture modeling in which volumetric models are the largest groups in process. Volumetric models are meant to describe 3D visual appearance of hand and arms. These models are used for analysis by synthesis tracking and in the identification of a body posture. Structures like cylinders and super quadrics, which are combination of simple spheres; circles hyper rectangles are used to shape body parts like forearms or upper arms. In 3D arm models the hand gesture is taken through a video input. The image is processed and it is compared with the model programmed in memory by calculating the various parameters and is then projected in 2d image. If both the images are similar the gesture is recognised and instruction is executed [2],[5].

2.1.2.2 *Appearance–based Models*

A large number of models belong to this group. These models take the outline of an object performing the gesture. The hand gesture is taken as a video input and then the 2D image is compared with predefined models. It checks for the skin colour matching which is then calibrated. This is then followed by the finger and contour detection and this pattern is matched. If both the images are similar the gesture is determined and executed [2],[5].

2.2 *Gesture Analysis*

It is the second phase of the gesture recognition system. It consists of two steps: Feature Detection and Parameter Estimation [5].

2.2.1 *Feature Detection*

Feature detection is a process where a person performing the gestural instructions is identified and the features to understand the instructions are extracted from the rest of the image. There are two types of cues: Color cues and Motion cues [5].

2.2.1.1 *Color Cues*

Colour cues take help of the characteristic colour footprints of the human skin. The colour footprints should be more distinctive and less sensitive to illumination changes. The drawback of this technique is the fluctuating of skin colour in varying lighting environments, which results in undetected, skin regions or wrongly detected non-skin textures. Common solution is to use restrictive backgrounds and clothing like uniform black backgrounds and dark long sleeves with the use of uniquely coloured gloves, which make it possible to detect the hand clearly in real time [5].

2.2.1.2 *Motion Cues*

Motion cues take into consideration certain assumptions about the gesturer. The assumptions may be that only one person gestures at a given point of time. Secondly, the gesturer should be stationary with respect to the background. The motion in the visual image is the movement of the arm of the gesturer whose movements are located, identified and executed. The drawback is when the background is not stationary [5].

2.2.1.3 *Overcoming the Problems of Cues*

- One way is the fusion of colour, motion and other visual cue or the fusion of visual cues with non-visual cues like speech or gaze [5].
- The second solution is to make use of prediction techniques. These techniques determine the estimates of the future feature locations based on the model dynamics and previous locations [5].

2.2.2 *Features and Detection*

Building *motion energy images* (MEI): MEIs highlight regions in the image where any form of motion was present. MEIs are 2D images which highlight the information of the motion of a sequence of 2D images by accumulating the motion of some characteristics image points [5]. Fingertip locations help obtain parameters of both, the 3D hand models and the 2D appearance-based gestural models. Fingertip locations can be obtained using marked gloves or colour markers to identify the fingertips [5]. Multiple cameras may be used to prevent one or more fingers being blocked by the palm from a given camera viewpoint and direction. However, restrictions are places on the user to posture his/her hand so that the occlusions are minimized [5].

2.2.3 *Parameter Estimation*

This is the last step of the gesture analysis stage. The type of computation depends on the model parameters and the features detected [5].

2.3 *Gesture Recognition*

It is the phase in which data analyzed from visual images of gestures is recognized as specific gesture [5].

Two tasks are commonly associated with the recognition process:

1) Optimal partitioning of the parameter space: Related to the choice of gestural models and their parameters [5].
2) Implementation of the recognition procedure: Key concern is the computational efficiency [5].

Gestural actions as opposed to static gestures involve both temporal and spatial context. A successful recognition scheme should consider the time space context of any specific gesture [5].

3. Classification Methods Of Hand Gestures

Thierry Messer, in [9] describes static hand gestures which are recognised using well-defined signs based on the posture of the hand. In a process, which is generally known and referred to as static hand gesture recognition, a person instructs the machine using his bare hands, whereas images of the persons hand gestures are captured and analyzed in order to determine the meaning of the hand gesture.

This paper also explains that the image of the hand is first captured using a system of two or more cameras [9]. This is followed by preprocessing stage which is used to optimally prepare the image obtained from the previous phase in order to extract the features [9]. The next stage is the feature extraction stage in which:

- Simplest method is to detect the hand’s outline which can be extracted using an edge detection algorithm.
- The other method described is Zernike Moments used to describe shapes by dividing the hand into two sub-regions – the palm and the fingers.
- The other idea is of local orientation histograms which consist of creating overlap-ping sub-windows, whereas each sub-window contains at least one pixel which lies inside the hand shape. For each of the sub-windows an orientation histogram is created, which is then added to the feature vector.
- The final method is using multi scale color features. Multi scale color features do not require any preprocessing of the image. It was proposed to perform the feature extraction directly in the color space, as this allows the combination of probabilistic skin-colors directly in the extraction phase. The advantage of directly working on a color image lies in the better distinction of hand and background regions [9].

The classification represents the task of assigning a feature vector or a set of features to some predefined classes in order to recognize the hand gesture. a class is defined as a set of reference features that were obtained during the training phase of the system or by manual feature extraction, using a set of training images. Therefore, the classification mainly consists of finding the best matching reference features for the features extracted in the previous phase [9].

- k-Nearest Neighbors: This classification method uses the feature-vectors gathered in the training to and the k nearest neighbors in an n-dimensional space. The training mainly consists of the extraction of (possible good discriminable) features from training images, which are then stored for later classification [9].

- Hidden Markov Models: The Hidden Markov Model (HMM) classifiers belong to the class of trainable classifiers. It represents a statistical model, in which the most probable matching gesture-class is determined for a given feature vector-based on the training data [9].

- Multi Layer Perceptron: A Multi Layer Perceptron (MLP) classifier is based on a neural network. Therefore, MLPs represent a trainable classifier (similar to Hidden Markov Models) [9].

4. Static Hand Gesture Recognition Software

Jan Kapoun, in his thesis [10] demonstrates his software application with the identification of static hand gestures in a real-time video feed. Jan describes the various techniques he has implemented for background segmentation and image recognition in his application. He also describes the shortcomings of his implementation. His project implementations were inspired from the Optical Character Recognition (OCR) model.

Jan has utilized the OpenCV library for implementing his software. OpenCV is a cross-platform computer vision library developed for real-time image processing. OpenCV contains all the algorithms Jan has used for this application like background subtraction, contours Hu Moments.

Jan utilized the concept that in order for a computer to recognize something in the screen, it is required to describe the object that needs to be recognized and it needs to be described in mathematical terms. Jan has described the following steps to implement visual recognition [10]:

- Scan the image of interest
- Subtract the background (i.e. everything we are not interested in)
- Find and store contours of the given objects
- Represent the contours in a given mathematical way.
- Compare the simplified mathematical representation to a pattern stored in computer's memory.

In order to mathematically represent an object for recognition, the object's contours are needed which can then turn into a row of numbers that can be subsequently stored in memory or further computed with. For this purpose, Jan has used the Freeman chain code.

Jan has also explained feature extraction based on colours, on rapid boost classifiers, and on background segmentation. Feature extraction is used to find out the location of the hand in the video feed. Feature extraction based on colours is used to identify the hand with its colour differentiated with the background. In feature extraction based on background segmentation, the computer has to learn the static background (all the objects that are not moving) and subtract it from the video feed.

Jan implements contour finding using OpenCV's functions. Contours represent a simplified version of a real-video image that can be compared against some pattern stored in memory. The solution used in Jan's application is a slight variation: first, he created the contours of a hand during a live-video feed and, in the real-time, computed the Hu moments of these contours. Subsequently, he stored these Hu moments in memory forfuture reference and comparison [10].

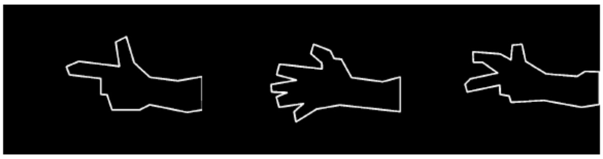

Fig. 1 – Contours of the hand gesture [10]

Jan also utilized a bounding rectangle which can be seen in the figure in order to make it easier for the software to know the location of the hand. In order to make the system recognize the hand, Jan places the hand in the bounding rectangle. Also once a hand is detected in the area of the bounded rectangle, a contour is visible around it.

The following figure (Fig 2) shows the layout of his application. To the left, the upper window shows the gesture recognized, and the bottom window shows the gestures in his database. In the center is Jan and his hand gesture in the bounding rectangle. To the right, the upper window shows the hand segmented from the background, and the lower window gives the mathematical values of the contours from the gesture.

Fig. 2 – Jan Kapoun's Application's Layout [10]

5. Security Elevator Using Gesture & Face Recognition

This application uses the concept of pattern recognition. The number of clusters depends on the number of patterns defined in the application. Usually, the less clusters cause higher recognition rate. If the number of clusters can be dynamically reduced, the overall recognition rate can be enhanced. The basic pattern recognition technique is to derive several liner/nonlinear lines to separate the feature space into multiple clusters [6].

In this elevator, there is no floor button for pressing. The decision for bringing someone to some floor is taken depends on his face and his hand gesture. The hand gesture indicates the floor he want to reach, and his face is used to decide that is this person permitted to reach the floor indicated or not based on his hand gesture [6].

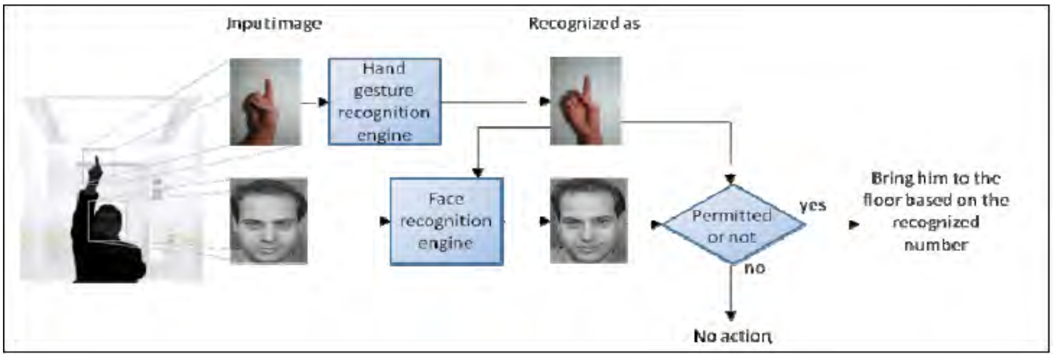

Fig. 3 – Security Elevator Algorithm [6]

- In this system, the input hand gesture and face images are extracted first.
- After that, the hand gesture and face recognition engine processes these two images partially simultaneously.
- After the hand gesture recognition result is produced, the face recognition engine eliminates the impossible candidates based on the recognized hand gesture dynamically, and figure out which is the recognized person.
- Finally, it is checked that is this person permitted to reach the floor indicated or not based on his hand gesture. If he is permitted, then the elevator will bring him to this floor, otherwise the elevator takes no action [6].

6. Sixth Sense Technology

When we come across something or someone we use our five senses to recognize or find some information about it. This senses do not help us in making the right decisions which easily accessible online. Sixth Sense as seen in Fig. 4 is a wearable gestural interface that supplements the physical world around us with digital information and lets us use natural hand gestures to interact with that information. It enables us to link with the digital information available using natural hand gestures. It automatically recognises the objects and retrieves information related to it. For example, for books from amazon it allows users to access it very easily fluently and is very user friendly. Sixth sense has the potential to be one of the best user interface that helps in accessing the information available online anywhere and anytime. Sixth sense is comprised of pocket projector mirror and a camera. The projector and the camera are connected to a mobile wearable device [7].

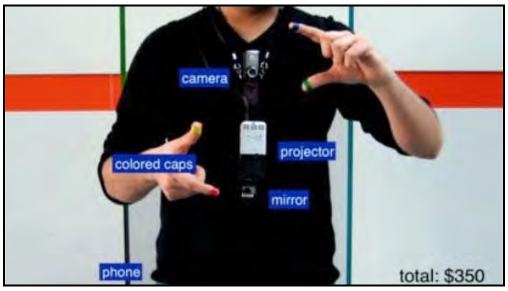

Fig. 4 – Components of Sixth Sense device [7]

6.1 Uses

6.1.1 *Calling*

Sixth sense prototype projects a keypad onto your hand and uses it to make a call [7].

6.1.2 *Time details*

User can draw a circle on your wrist to get virtual watch that gives you the correct time [7].

6.1.3 *Access book information*

System can project amazon ratings on the book as well as reviews and other relevant information [7].

6.1.4 *Take pictures*

On can fashion his index fingers and thumbs into a square (framing gesture), the system will snap a photo [7].

7. Conclusion And Future Scope

Human Computer Interaction is still in its infancy. Visual interpretation of hand gestures today allows the development of potentially natural interfaces to computer-controlled environments. Though most current systems employ hand gestures for manipulation of objects the complexity of the interpretation of gestures dictates the achievable solution. Hand gestures for HCI are mostly restricted to single handed and produced by single user in the system. This consequently downgrades the effectiveness of the interaction. Computer Vision methods for hand gesture interfaces must surpass current performance in terms of robustness and speed to achieve interactivity and usability. Considering the relative infancy of research related to vision based gesture recognition remarkable progress has been made. To continue this momentum it is clear that further research in the areas of feature extraction, classification methods and gesture representation are required to realize the ultimate goal of humans interfacing with machines on their own natural terms.

References

[1] Margaret Rouse, "Definition: Gesture Recognition,". [Online]. Available: http://whatis.techtarget.com/definition/gesture-recognition. [Accessed: Sept. 16, 2012].
[2] Pragati Garg, Naveen Agarwal and Sanjeev Sofat, "Vision Based Hand Gesture Recognition," World Academy of Science, Engineering and Technology, 49, pp. 972-977,2009.
[3] Bill Buxton and Mark Billinghurst, "Gesture Based Interaction," Haptic Input, pp. 14.1-14.28,May 27, 2009.
[4] Mahafreed Irani, "Enables For the Disabled," The Times of India, Mumbai: Times Review-Techtonic, pp. 24, Jan 23, 2011.
[5] Vladimir, 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, pp.677-692, July 1997.
[6] Yung-Wei Kao, Hui-Zhen Gu, and Shyan-Ming Yuan, "Integration of face and hand gesture recognition," Third 2008 International Conference on Convergence and Hybrid Information Technology, pg. 330-332, Nov 10, 2009.
[7] Pranav Mistry, "Sixth Sense," Sixth Sense: A wearable gestural interface. [Online]. Available: http://www.pranavmistry.com/projects/sixthsense. [Accessed: Sept. 10, 2012].
[8] David Geer, "Will Gesture- Recognition Technology Point the Way?," in Computer, IEEE Computer Society, vol. 37, no. 10, pp. 20-23, Oct. 2004.
[9] Thierry Messer, "Static hand gesture recognition," University of Fribourg, Switzerland, 2006.
[10] Mgr. Jan Kapoun, "Static Hand Gesture Recognition Software," Bachelor's Thesis, Univ. of South Bohemia, Informatics Dept., České Budějovice, 2010.