

hand gesture recognition

srikanth

nageswara rao

jaya venkatesh

Abstract

Vision based real time gesture recognition system received a great attention in recent years because of its manifoldness applications and the ability to interact with system efficiently through human computer interaction. In this paper, a review of recent hand gesture recognition systems is presented. This paper includes a brief review on camera interface. Image processing, hand gestures, colour detection and hand gesture recognition. Advantages and drawbacks of the system are mentioned finally

Introduction

The essential aim of building hand gesture recognition system is to create a natural interaction between human and computer where the recognition gestures can be used for controlling a robot or conveying meaningful information [1]. The aim of this technique is the proposal of a real time vision system for its application within visual interaction environments through hand gesture recognition, using general purpose hardware and low cost sensors, like a simple personal computer and an USB Webcam, so any user could make use of it in his/her office or home [2].

The development in computer vision makes it possible to approach the interface problems from a human perspective, making the communication between computer and humans more natural [6]. When we as humans communicate, we use our voice and parts of our body such as our face and arms in making gestures. Recently, many attempts have been made to create system that through computer vision could be able to understand gestures [6].

There are two main characteristics that should be deemed when designing a HCI system as mentioned in [3] : Functionality and Usability. System functionality referred to the set of functions or services that the system equips to the users [3]

while system usability referred to the level and scope that the system can operate and perform specific user purposes efficiently [3].

The system that attains a suitable balance between these concepts considered as influential performance and powerful system [3]

Compared to many existing interfaces, hand gestures have the advantages of being easy to use, natural and intuitive. Successful applications of hand gesture recognition include computer game control, human robot interaction and sign language recognition to name a few [4].

The research on visual interpretation of hand gestures for Human-computer Interaction has been a topic of research for many years. The main attempt of the HCI is to communicate naturally with the machines. On similar lines, there is a tremendous progress in speech recognition.

To exploit the use of gestures in HCI, it is necessary to provide the means by which they can be interpreted by the computers [7]. The HCI requires a static/dynamic movements of human hand as an input. Apart from human hand, HCI can pick up the gesture by the movement of human arm, face or even other parts of the human body. Initially, attempts for this HCI using hand gestures required to use hand gloves. These were known as Glove-Based Devices [7][8][9][10]. Glove based gestural interfaces require the user to wear a cumbersome device and generally carry a load of cables that connect the device to a computer. This hinders the ease and naturalness with which the user can interact with the computer controlled environment[7].

Eventually the awkwardness and unnaturalness of using gloves was overcome with an advent in HCI where the video based non contact interaction techniques required a video cameras or the computer visions as an input device which would take human hand gestures as an input. These vision based non contact interaction made the gesture inputs more natural.

2. System components

A low cost computer vision system that can be executed in a common personal computer equipped with an USB webcam is one of the main component of the system. The system should be able to work under different degrees of scene background complexity and illumination conditions.

3. Hand gestures

Outside the HCI framework, hand gestures cannot be easily defined. The definition if they exist, are particularly related to communicational aspect of human hand and body movement. 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 of 2. Interacting with the environment[11]. However in the domain of HCI, the notion of gestures is somewhat different. In a computer controlled environment, one wants to use the human hand to perform tasks that mimic both the natural use of the hand as a manipulation and its use in human machine communication (Control of Computer/Machine functions through gestures.)[12]

All hand movements are classified into two major classes:

- ⌚ Gestures
- ⌚ Unintentional Movements

The unintentional movements must be differentiated from the gestures to avoid the meaningless output from the system as the unintentional movements do not convey meaningful information. [7]

Gestures can be static (the users assume a certain pose or configuration) or dynamic. Some gestures also have both static and dynamic elements as in sign language.[11]

3.1. Modelling of gesture

A more appealing approach suitable to real time computer vision lies in the use of simple 3D geometric structures to model in human body [13]. The 3D hand and arm models have often been a choice for hand gesture modelling. They can be classified in two large groups:

- ⌚ Volumetric Models
- ⌚ Skeletal Models

Volumetric models are meant to describe the 3D visual appearance of the human hand and arms. They are commonly found in the field of computer animation but recently also have been used in computer vision application.[14]

While static hand gestures are modelled in terms of hand configuration, as defined by the flex angles of the fingers and palm orientation, dynamic hand gestures include hand trajectories and orientation in addition to these. So, appropriate interpretation of dynamic gestures on the basis of hand movement in addition to shape and position is necessary for recognition [5]

4. Image preprocessing

Hand gesture inputs is taken from the web cam and stored in visual memory which is created in a start up step.

A frame from the web cam is captured and each frame is processed separately before its analysis [2]. Two generally sequential tasks are involved in the analysis. The first task involves —detecting or extracting relevant image features from raw image or image sequence. The second task uses these image features for computing the model parameters [7]. In detection process, it is first necessary to localize the gesturer. Once the gesturer is localized; the desired set of features can be detected.

In dynamic hand gesture recognition, after tracking the hand motion from the gesture video sequence, subsequently, trajectory is estimated through which the hand moves during gesticulation [5]. While trajectory estimation is quite simple and straight forward in Glove based hand gesture recognition system [16] [17] that provide spatial information directly, trajectory estimation in vision based system may require to apply complex algorithm to track hand and fingers using silhouettes and edges.

4.1. Localization

Gesture localization is a process in which the person who is performing the gestures is extracted from the rest of the visual image.

.

4.2. Segmentation

Segmentation is the next process for recognizing hand gestures. It is the process of dividing the input image (in this case hand gesture) into regions separated by boundaries. The segmentation process depends on the type of gesture if it is dynamic gesture then the hand gesture need to be located and tracked, if it is static gesture (posture), the input image have to be segmented only.

5. Colour detection

[18] Skin colour detection plays an important role in a wide range of image processing applications ranging from face detection, face tracking, gesture analysis, content-based image retrieval (CBIR) systems and to various human computer interaction

domains. The common helpful cue used segmenting the hand is the skin colour. Skin colour detection in visible spectrum can be a very challenging task as the skin colour in an image is sensitive to various factors such as :

- ⌚ Illumination
- ⌚ Camera Characteristics
- ⌚ Ethnicity
- ⌚ Individual Characteristics of the gesturer
- ⌚ Other factors like background colours, shadows & motions also influence skin colour appearance.

The major drawback of colour based localization techniques is the variability of the skin colour footprint in different lighting conditions. This frequently results in undetected skin regions or falsely detected non skin textures. The common solution to this problem is the use of restrictive backgrounds and clothing [7].

It has been demonstrated [19] that regardless of ethnicity, the chrominance of skin is quiet consistent. The major difference between skin tones is intensity, as dark skin people have greater skin saturation then light skin people [20].

[18] The HSV space defines colour as hue, saturation and brightness (Lightness or value). The transformation of RGB to HSV is invariant to high

intensity at white lights, ambient light and surface orientations relative to the light source and hence can form a very good choice for skin detection methods. [20] However, the YUV colour system is employed for separating chrominance and intensity. The symbol

'Y' denotes intensity while the 'UV' components specify chrominance components. Human skin colours are distributed over a very small region in the chrominance plane.

6. Blobs analysis

Blobs, binary linked objects are groups of pixels that share the same label due to their connectivity in a binary image [2]. After a blob analysis, all those pixels that belong to a same object share a unique label so that every blob can be identified with a label.

7. Gesture recognition

Good segmentation process needs to leads to perfect feature extraction process and the later plays an important role in a successful recognition process [3]. Gesture recognition is an ideal example of multidisciplinary research. There are different tools for gesture recognition, based on approaches ranging from statistical modelling, computer vision and pattern recognition, image processing etc. Most of the problems have been addressed based on statistical model such as HMM's ,ANN's etc[18]. We are going to see some learning algorithms in brief

7.1 Learning Algorithms

Here we describe the use of three of the most common learning algorithms used to recognize hand gestures. These algorithms all stem from the artificial intelligence community, and their common trait is that recognition accuracy can be increased through training.[24]

7.1.1 Neural Networks

This section presents a brief introduction into the concepts involved in neural networks.[25] A neural network is an information processing system loosely based on the operation of neurons in the brain. While the neuron acts as the fundamental functional unit of the brain, the neural network uses the node as its fundamental unit; the nodes are connected by links, and the links have an associated weight that can act as a storage mechanism. Each node is considered a single computational unit containing two components. The first component is the input function which computes the weighted sum of its input values; the second is the activation function, which transforms the weighted sum into a final output value. Many different activation functions can be used; the step, sign, and sigmoid functions being quite common. Since they are all relatively simple to use. For example, using the step function, if the weighted sum is above a certain threshold, the function outputs a one indicating the node has —fired otherwise it outputs a zero indicating the node has not fired. The other two activation functions act in a similar manner. Neural networks generally have two basic structures or topologies, a feed-forward structure and a recurrent structure. A feed-forward network can be considered a directed acyclic graph, while a recurrent network has an arbitrary topology. The recurrent network has the advantage over a feed-forward network in that it can model systems with state transitions Training is an important issue in neural networks and can be classified in two different ways. First, supervised learning trains the network by providing matching input and output patterns; this trains the network in advance and as a result the network does not learn while it is running. The second learning mechanism is unsupervised learning or self-organization which trains the network to respond to clusters of patterns within the input. There is no training in advance and the system must develop its own representation of the input, since no matching output is provided. Neural networks are a useful method for recognizing gestures, yield increased accuracy conditioned upon network training, and work for both glove based and vision-based solutions. However, they have distinct disadvantages. First, different configurations of a given network can give very different results,

Training is an important issue in neural networks and can be classified in two different ways. First, supervised learning trains the network by providing matching input and output patterns; this trains the network in advance and as a result the network does not learn while it is running. The second learning mechanism is unsupervised learning or self-organization which trains the network to respond to clusters of patterns within the input. There is no training in advance and the system must develop its own representation of the input, since no matching output is provided. Neural networks are a useful method for recognizing gestures, yield increased accuracy conditioned upon network training, and work for both glove based and vision-based solutions. However, they have distinct disadvantages. First, different configurations of a given network can give very different results, and it is difficult to determine which configuration is best

without implementing them. Another disadvantage is the considerable time involved in training the network. Finally, the whole network must be retrained in order to incorporate a new gesture. If the gesture set is known beforehand this is not an issue, but if gestures are likely to change dynamically as the system develops, a neural network is probably not appropriate.

Strengths:

1. Can be used in either a vision- or glove-based solution
2. Can recognize large posture or gesture sets
3. With adequate training, high accuracy can be achieved

Weaknesses:

1. Network training can be very time consuming and does not guarantee good results
2. Requires retraining of the entire network if hand postures or gestures are added or removed.

7.1.2 Hidden Markov models

In the Markov model, the state sequence is observable. The output observable event in any given state is deterministic, not random. This will be too constraining when we use it to model the stochastic nature of the human performance,

which is related to doubly stochastic processes, namely human mental states (hidden) and human actions (observable). It is necessary that the Observable event is a probabilistic function of the HMM is a representation of a Markov process and is a doubly embedded stochastic process with an underlying stochastic process that cannot be directly observed, but can only be observed through another set of stochastic processes that produce the sequence of observable symbol

7.2. Model evaluation process

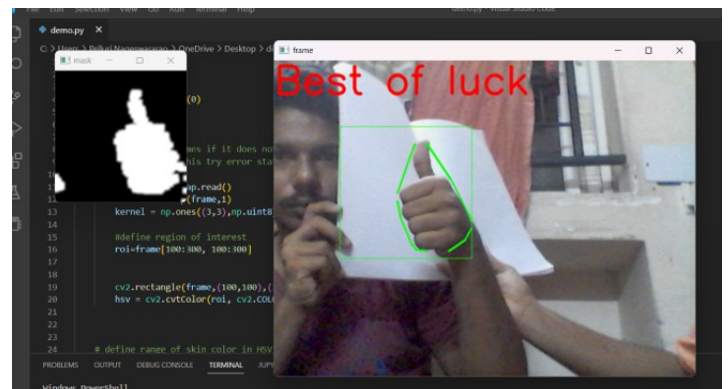
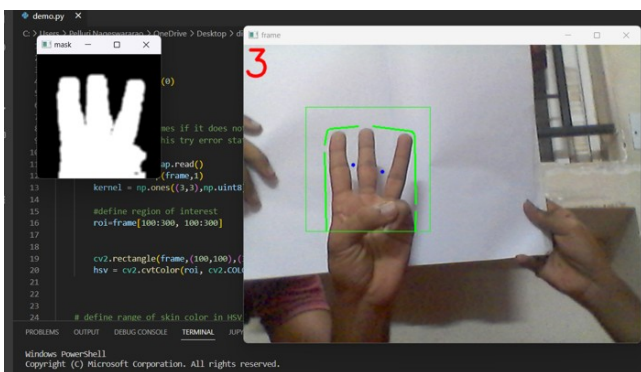
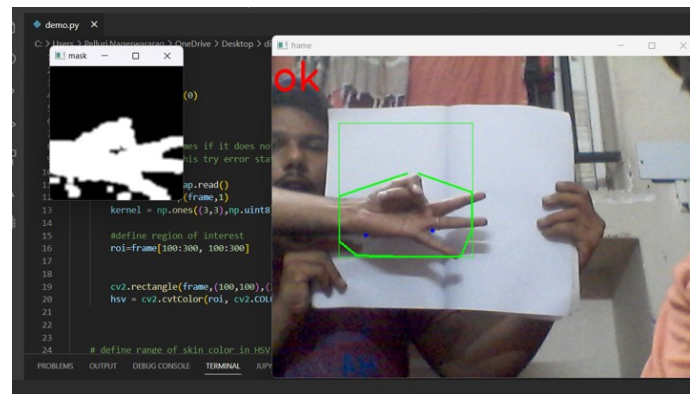
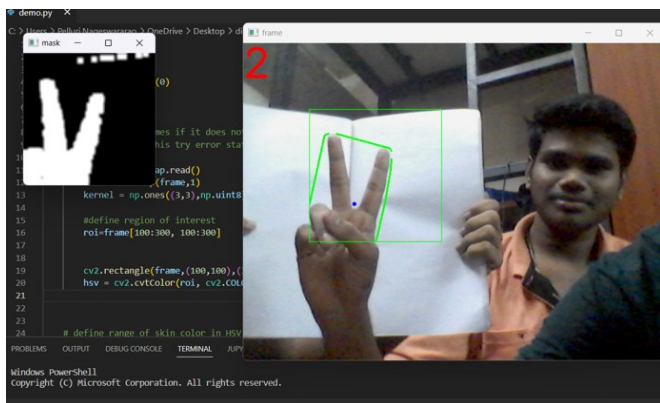
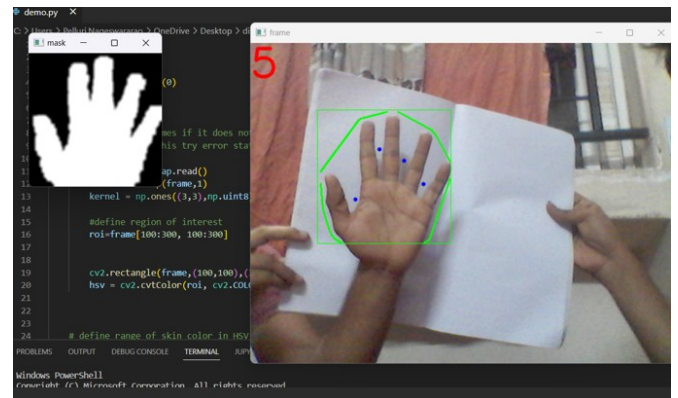
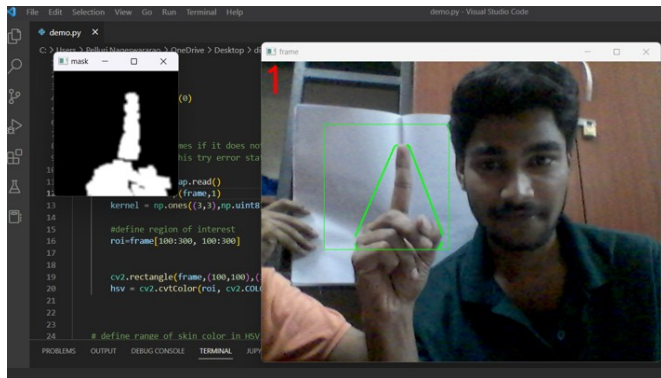
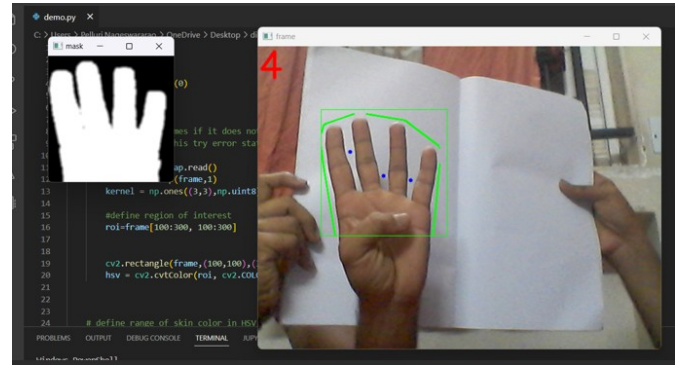
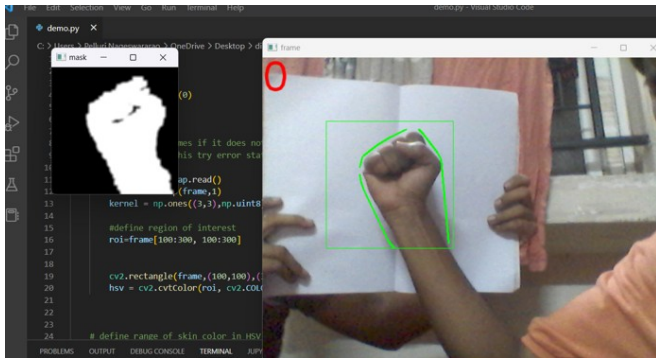
In this process, a gesture trajectory is tested over the set of trained HMM's in order to decide which gesture it belongs to from the database, however, the larger the number of trained HMM's (gestures), the more the computationally demanding the recognition Procedure

8. Applications of hand gesture

Gesture recognition has wide range in applications such as following: [18]

- ⌚ Developing aids for the hearing impaired
- ⌚ Enabling very young children to interact with computer
- ⌚ Designing techniques for forensic identification
- ⌚ Recognizing sign language
- ⌚ Medically monitoring patients
- ⌚ Navigating and/or manipulating virtual environments
- ⌚ Communicating in video conferencing
- ⌚ Distance learning / tele-teaching assistance
- ⌚ Graphic editor control

9.OUTPUTS:



10. Conclusion and discussion

This survey has provided a comprehensive overview of hand gesture recognition techniques. Hand gestures are an interesting interaction insight in a variety of computer applications. Two principal questions must be answered when using them: The first question is what technology to use for collecting raw data from the hand. Generally, two types of technologies are available for collecting this raw data. The first one is a glove input device and the second way of collecting raw data is to use computer vision. Accuracy of a glove input device depends on the type of sensor technology used; usually, the more accurate the glove is, the more expensive it is. In a vision-based solution, one or more cameras placed in the environment record hand movement. Both types of solutions have their own advantages and disadvantages, and the question that which solution to use is a difficult one. However, when using a hand posture or gesture-based interface, the user does not want to wear the device and be physically attached to the computer. If visionbased solutions can overcome some of their difficulties and disadvantages, they appear to be the best choice for raw data collection. The second question to be answered when using hand gestures is what recognition technique will maximize accuracy and robustness. A number of recognition techniques are available and in some cases, where hand gesture recognition is to be used, the possibilities are narrowed down, as some algorithms are best fitted only for gesture recognition. This review has categorized these techniques into three broad categories: Feature extraction, statistics, and model learning algorithms.

There are a number of interesting areas for future research in hand gesture recognition. The field is not yet very matured – we have a long way to go before this type of concept is robust enough to be seen in commercial, main stream applications. Research into

better hardware for data collection is important and HMM technique for gesture recognition was found to be more accurate compared to neural networks and instance based model. This survey has provided a comprehensive overview of hand gesture recognition techniques.