Analysis of Vision Based Hand Gesture Recognition Systems

Abhishek R Shetty¹, Arjun V², Bharath R³, Goutham Ramesh⁴
B.E. Final Year Students⁴
Department of Computer Science & Engineering
BMS Institute Of Technology, Bangalore-64

ABSTRACT

Vision based real time gesture recognition systems have received a great amount of attention in recent years because of its manifold applications and the ability to interact with system efficiently through human computer interaction. In this paper, a review of recent vision based hand gesture recognition systems is presented. This paper includes a brief review on various vision based approaches available. Advantages and drawbacks of the systems are mentioned finally.

Keywords - Vision based hand gesture recognition; HCI; Detection and segmentation; Classification;

Corresponding Author: Vishakha Yadav, Assistant Professor, Dept. Of CSE,BMSIT.

1. INTRODUCTION

Communication in daily life is performed via the help of vocal sounds and body language. However vocal sounds are the main tool for interaction, body language and facial expressions have a serious support in the meanwhile. Even in a few cases, interacting with the physical world by using those expressive movements instead of speaking is much easier. Body language has wide range of activities namely eye expressions, slight change in skin color, variation of the vibrations in vocal sounds etc. But the most important body language expressions are performed mainly using hands. Hand gestures would be very helpful for exchanging information in recent cases such as pointing out an object located at a random place, representing a number, expressing a feeling etc. and also hand gestures are the primary interaction tools for sign language [1] and gesture based computer control.

With the help of serious improvements in the image acquisition and processing technology, hand gestures become a significant and popular tool in human machine
interaction (HCI) systems. Recently, human machine interfaces are based on and limited to use keyboards and mice with some additional tools such as special pens and touch screens [2]. Main application areas of hand gesture recognition in human machine interface systems are keyboard-mouse simulations, special game play without joysticks, sign language recognition, 3D animations, motion and performance capture system, special HCI for disabled users etc.

Non-vision based approach are fast growing techniques that require modifications of the environment, e.g. that the users wear gloves of particular colour distribution or with particular markers [3].

Vision based approach are techniques that are based on cameras operating beyond the visible spectrum. Vision has the potential of carrying a wealth of information in a non-intrusive manner and at a considerably low cost, therefore it constitutes a very attractive sensing modality for developing perceptive user interfaces.

2. VISION BASED ALGORITHMS

Hand gesture recognition using vision based approach is an important aspect in Human-Computer interaction, and can be used in various fields, such as virtual reality, sign language recognition, computer games and also plays a major role in Human-Computer Intelligent interaction which is a combination of many interdisciplinary studies. Vision has the potential of carrying a wealth of information in a non-intrusive.

Some of the challenges faced by vision based hand gesture recognition are as follows-

• The hand gesture should be recognised accurately, and the interaction should response correctly. Then the solution should run flawlessly, and responses in real-time.

• Then the solution should run flawlessly, and respond in real-time.

• The tracking, recognition and interaction algorithm should have good speed performance.

• The solution should be robust to many defects on the vision based sensors including image noise and lighting conditions and motion blur.

• Finally, the solution should not have too much constrains, for example, it should be workable at various resolutions that equals to allowing a degree of freedom on the distance between hand and the sensor.

To solve vision-based hand gesture recognition and interaction problem, it is necessary to encounter multiple steps-

• Initially, it is necessary to detect and locate the hand position.

• Further, hand is isolated from the background noise by doing the hand segmentation.
• Next thing to do is to track the hand and estimate its representations with certain features, getting the hand feature description.

• Then a data description of the extracted feature is used to do the recognition through classification or other methods.

• At last, an interaction implementation should be proposed.

3. VARIOUS APPROACHES

The vision based methods are broadly classified into three major types namely-

• 3D based model
• Appearance based model
• Learning based model

3.1 3D based

Here, we use 3D representation of the images which are obtained using special depth sensitive or multiple cameras. The 3D model approach is further subdivided into –

• Volumetric based model
• Skeletal based model

3.1.1 Volumetric based model

Volumetric approaches have been heavily used in computer animation industry and for computer vision purposes. These models are generally created of complicated 3D surfaces, like NURBS [4] or polygon meshes [5].

The drawback of this method is that is very complex, and systems for live analysis are still to be developed. At the moment, a more interesting approach would be to map simple primitive objects to the person’s most important body parts (for example cylinders for the arms and neck, sphere for the head) and analyze the way these interact with each other.

3.1.2 Skeletal based model

Instead of using intensive processing of the 3D models and dealing with a lot of parameters, we can just use a simplified version of joint angle parameters along with segment lengths. It is known as a skeletal representation of the body, where a virtual skeleton [6] of the person is computed and parts of the body are mapped to certain segments. The analysis here was done using the position and orientation of these segments and the relation between each one of them (for example the angle between the joints and the relative position or orientation)

Advantages of using skeletal models:

• Algorithms are faster because only key or important parameters are analyzed.
• Pattern matching [7] against a template database is possible
• Using key or important points allows the detection program to focus on the significant parts of the body
3.2 Appearance Based

These models don’t use a spatial representation of the body anymore, because they derive or obtain the parameters directly from the images or videos using a template database.

They are further subdivided into-

• Deformable 2D Templates
• Image Sequences

3.2.1 Deformable 2D Templates

Deformable 2D templates[8] are the sets of points on the outline or circumference of an object, is used as interpolation nodes for the object outline approximation. The simplest interpolation function that’s used is a piecewise linear function. The templates consist of the average point sets, point variability parameters, and so-called external deformations. Average point sets describe the “average” shape within a certain group of shapes.

3.2.2 Image Sequences

Human gesture recognition in image sequences [9] has many applications including human-computer interaction, surveillance, and video games. Parameters for this method are either the images themselves, or certain features derived from these images. Most of the time, only one (monoscopic) or two (stereoscopic) views[10] are used. Hand gesture inputs is taken from the web cam and stored in visual memory which is created in a startup step. A frame from the web cam is captured and each frame is processed separately before its analysis. 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. 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. While trajectory estimation is quite simple and straight forward in Glove based hand gesture recognition system 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[11].

3.3 Learning based methods

These algorithms all stem from the artificial intelligence community, and their common attribute is that recognition accuracy can be increased through training. They are further classified into –

• Instance based learning
• Neural Networks

3.3.1 Instance based Learning

Instance-based learning is another recognition technique that stems from work done in machine learning. The main difference between instance-based learning and other learning algorithms such as neural networks and hidden Markov models [13] is the way in which the training data is used. With supervised neural networks, for example the training data is passed through the network and the weights at various nodes are updated to fit the training set. With
instance-based learning, the training data is simply used as a database in which to classify other instances. An instance, in general, is a vector of features of the entity to be classified. For instance, in posture and gesture recognition, a feature vector might be the position and orientation of the hand and the bend values for each of the fingers. Instance-based learning methods include techniques that represent instances as points in Euclidean space, such as the K-Nearest Neighbor algorithm [15], and techniques in which instances have a more symbolic representation, such as case-based reasoning [12]. In the K-Nearest Neighbor algorithm, an instance is a feature vector of size n with points in n-dimensional space.

Another type of instance-based learning technique is case-based reasoning, in which instances have more elaborate descriptions. Instance-based learning techniques have the advantage of simplicity, but they have a number of disadvantages as well. One major disadvantage is the cost of classifying new instances. Another disadvantage of these methods is that not all of the training examples may fit in main memory, and thus will also increase response time. Unfortunately, very little work has been done on instance-based learning in recognizing hand postures and gestures. More research is needed to determine whether the technique can be applied to hand gestures and if the accuracy can be improved.

Strengths
1. Except for case-based reasoning, instance-based learning techniques are relatively simple to implement.
2. Can recognize a large set of hand postures with moderately high accuracy.
3. Provides continuous training.

Weaknesses
1. Requires a large amount of primary memory as training set increases.
2. Response time issues may arise due to a large amount of computation at instance classification time.
3. Only a little reported in the literature on using instance-based learning with hand postures and gestures.

3.3.2 Neural Networks

This section presents a brief introduction into the concepts involved in neural networks. A neural network [14] 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 has 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 may be used; the step, sign, and sigmoid functions are quite common. Since they are all relatively simple to use. For example, making use of the step function, if the weighted sum is above a certain threshold or limit, 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. The feed-forward network can be considered as a directed acyclic graph, while a recurrent network can be considered as an arbitrary topology. The recurrent network has the advantage over a feed-forward network [16] 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 an 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 inputs, since no matching outputs are provided.

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.
4. Applications and Drawbacks

Gesture recognition has wide range in applications such as following:
- Developing or creating aids for the hearing impaired.
- Enabling young children to interact with computer.
- Designing techniques or methods 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.

Although there are numerous advantages of dynamic hand gesture recognition system, there are a few drawbacks or limitations to it which are as follows:
- The number of cameras used.
- Their speed and latency.
- Structure of environment (Restrictions such as lighting or speed of movement)
- Any user requirements (Whether user must wear anything special)
- The low level features used (Edges, regions, Silhouettes moments, histograms)
- Whether 2D or 3D representation is used.
- Irrelevant objects might overlap with the hand.
- Recognition limited to numbers only.
REFERENCES

1. Rung-Huei Liang, Ming Ouhyoung, A Real-time Continuous Gesture Recognition System for Sign Language, Dept. of Information Management, Shih-Chien University, R.O.C. Dept. of Computer Science and Information Engineering, National Taiwan University.
2. Andrew Sears, Catherine Plaisant, Ben Schneiderman, A new era for touchscreen applications: High precision, dragging icons and refined feedback, Human computer interaction laboratory, Dept of Computer Science, University of Maryland.
3. Farid Parvini, Dennis McLeod, Cyrus Shahabi, Bahareh Navai, Baharak Zali, Shahram Ghandeharizadeh, An Approach to Glove-Based Gesture Recognition, Computer Science Department, University of Southern California.
6. Yuan Shi, Qi wei, Ruijie liu, Yuli Ge, Simplified Skeleton Model based Hand Gesture Recognition, Dalian Institute of Science and Technology.
9. Mark Bayazit, Alex Couture-Beil, Greg Mori, Real-time Motion-based Gesture Recognition using the GPU, School of Computing Science Simon Fraser University, Burnaby, BC, Canada.
11. Daniel Mohr, Gabriel Zachmann, University of Bremen, Germany, A Survey of Vision-Based Markerless Hand Tracking Approaches.
12. Dr. Ralph Bergmann, Prof. Michael. M. Richter, Introduction to Case-Based Reasoning Centre for Learning Systems and Applications Research Group, Department of Computer Science, University of Kaiserslautern.