REAL TIME MOTION CONTROLLING USING HAND GESTURE RECOGNITION

A PROJECT BY
CHIRAG AGARWAL
SOUVIK DAS
RUDROJIT CHANDA
OF
FUTURE INSTITUTE OF ENGINEERING AND MANAGEMENT
ABSTRACT

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. In this report a review of vision based hand gesture recognition is presented. The existing approaches are categorized into 3D model based approaches and appearance based approaches, highlighting their advantages and shortcomings and identifying the open issues. Gesture recognition can be seen as a way for computers to begin to understand human body language, thus building a richer bridge between machines and humans.
## CONTENTS

<table>
<thead>
<tr>
<th>Chapter</th>
<th>Page No.</th>
</tr>
</thead>
<tbody>
<tr>
<td>I.   Introduction</td>
<td>3</td>
</tr>
<tr>
<td>II.  System Overview</td>
<td>4</td>
</tr>
<tr>
<td>A. Pre-Processing</td>
<td>5</td>
</tr>
<tr>
<td>B. Recognition</td>
<td>6</td>
</tr>
<tr>
<td>C. Controlling</td>
<td>8</td>
</tr>
<tr>
<td>III. Gesture Based Applications</td>
<td>9</td>
</tr>
<tr>
<td>IV.  Advantages and Disadvantages</td>
<td>11</td>
</tr>
<tr>
<td>Conclusion</td>
<td>12</td>
</tr>
<tr>
<td>References</td>
<td>12</td>
</tr>
</tbody>
</table>
I. INTRODUCTION

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. The approaches present can be mainly divided into “Data-Glove based” and “Vision Based” approaches. The Data-Glove based methods use sensor devices for digitizing hand and finger motions into multi-parametric data. The extra sensors make it easy to collect hand configuration and movement. However, the devices are quite expensive and bring much cumbersome experience to the users. In contrast, the Vision Based methods require only a camera, thus realizing a natural interaction between humans and computers without the use of any extra devices. These systems tend to complement biological vision by describing artificial vision systems that are implemented in software and/or hardware. This poses a challenging problem as these systems need to be background invariant, lighting insensitive, person and camera independent to achieve real time performance. Moreover, such systems must be optimized to meet the requirements, including accuracy and robustness. 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.

Gesture recognition can be conducted with techniques from computer vision and image processing. 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. 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.

This project will design and build a man-machine interface using a video camera to interpret various geometrical shapes constructed in front of the camera using red colour bands worn on our fingers and controlling their respective 3-D models, viz. performing Scaling and Translation properties on the shapes, and we have constructed a simple 3-D game of placing the shapes onto their respective places within the allotted time period. The keyboard and mouse are currently the main interfaces between man and computer. In other areas where 3D information is required, such as computer games, robotics and design, other mechanical devices such as roller-balls, joysticks and data-gloves are used. Humans communicate mainly by vision and sound, therefore, a man-machine interface would be more intuitive if it made greater use of vision and audio recognition. Another advantage is that the user not only can communicate from a distance, but need have no physical contact with the computer. However, unlike audio commands, a visual system would be preferable in noisy environments or in situations where sound would cause a disturbance. The visual system chosen was the recognition of hand gestures. The amount of computation required to process hand gestures is much greater than that of the mechanical devices, however standard desktop computers are now quick enough to make this project — hand gesture recognition using computer vision — a viable proposition.
II. SYSTEM OVERVIEW

Vision based analysis, is based on the way human beings perceive information about their surroundings, yet it is probably the most difficult to implement in a satisfactory way. Several different approaches have been tested so far.

- One is to build a three-dimensional model of the human hand. The model is matched to images of the hand by one or more cameras, and parameters corresponding to palm orientation and joint angles are estimated. These parameters are then used to perform gesture classification.

- Second one to capture the image using a camera then extract some feature and those features are used as input in a classification algorithm for classification.

In this project we have used second method for modelling the system. The major operations performed by the system are summarized in the figure given below Fig. 1.

Fig. 1 System Architecture

Here we can see that live video is recorded with the use of a standard web-camera. The video is then being processed frame-wise and every frame of the input video is scanned for various 2-D geometrical figures formed by detecting the red colour bands worn on our fingers and hence are fed into the gesture recognition system, in which it is processed and compared efficiently with the help of an algorithm. Segmentation and morphological filtering techniques are applied on images in pre-processing phase then using contour detection we will obtain our prime feature. As soon as the algorithm detects the presence of a 2-D geometrical figure, the virtual object or the 3-d model is then being updated accordingly and the user interfaces with machine with the help of a user interface display.
A. PRE-PROCESSING

Pre-processing is very much required task to be done in hand gesture recognition system. Pre-processing is applied to images before we can extract features from each frame of hand images.

Pre-processing consists of two steps:

• Median filtering
• Segmentation
• Morphological opening

In the pre-processing stage first we break the whole video into its respective frames and then in order to detect red colour blobs (objects) from each frame we subtract the red component of the acquired frame from the gray scale version of the same acquired frame. This is a very important step as we have to adjust the threshold of the red colour in accordance with the background, so that the red colour detection is satisfactorily performed. After acquiring the same we perform a 3x3 median filtering operation in order to remove different noises obtained along with the image.

Now Segmentation is done to convert the previous obtained image into binary image so that we can have only two objects in image one is the detected red objects (WHITE) and other is background (BLACK). After all this some small objects still remain in the binary image. They are removed with the help of Morphological open operation.
B. RECOGNITION

In the previous section, methods were devised to obtain accurate information about the position of red colour band pixels. This information is then used to calculate the centroids of the red colour bands with subsequent data pertaining to hand movement and scaling. The next step is to use all of this information to recognise the gesture within the frame.

A. Choice of recognition strategy

Two methods present themselves by which a given gesture could be recognised from two dimensional ‘silhouette’ information:

**Direct method based on geometry:** Knowing that the hand is made up of bones of fixed width connected by joints which can only flex in certain directions and by limited angles it would be possible to calculate the silhouettes for a large number of hand gestures. Thus, it would be possible to take the silhouette information provided by the detection method and find the most likely gesture that corresponds to it by direct comparison. The advantages of this method are that it would require very little training and would be easy to extend to any number of gestures as required. However, the model for calculating the silhouette for any given gesture would be hard to construct and in order to attain a high degree of accuracy it would be necessary to model the effect of all light sources in the room on the shadows cast on the hand by itself.

**Learning method:** With this method the gesture set to be recognised would be “taught” to the system beforehand. Any given gesture could then be compared with the stored gestures and a match score calculated. The highest scoring gesture could then be displayed if its score was greater than some match quality threshold. The advantage of this system is that no prior information is required about the lighting conditions or the geometry of the hand for the system to work, as this information would be encoded into the system during training. The system would be faster than the above method if the gesture set was kept small. The disadvantage with this system is that each gesture would need to be trained at least once and for any degree of accuracy, several times. The gesture set is also likely to be user specific. It was decided to proceed with the learning method for reasons of computation speed and ease of implementation.

Of the above two methods we chose the direct method in which we detected the red colour bands worn on the fingers and with that were able to draw different geometrical figures and represented their respective 3-D model via a 3-D game.
B. Selection of test gesture set

We obtain the centroid and the bounding box of all the detected blobs (in our case the number of Red colour Objects), and once we are able to do that we can easily calculate the number of blobs by obtaining the length of the centroid matrix.

Once we have the required blobs along with their centroids and bounding boxes we can simply join all the centroid together and hence get the resultant geometrical figure.

Fig.6 The Red colour Band detection resulting in various Geometrical figures: (a) Quadrangle, (b) Triangle, (c) Circle
C. Controlling

Once we have recognised the different figures the next task we perform is to control their respective Scaling and Translation properties. These two properties if controlled can give rise to a new versatile world where we are able to control different modelling figures with our hand gestures and also control their respective areas and move them accordingly.

A. Scaling: In Euclidean geometry, Uniform Scaling is a linear transformation that enlarges (increases) or shrinks (diminishes) objects by a scale factor that is the same in all directions. The result of uniform scaling is similar (in the geometric sense) to the original. A scale factor of 1 is normally allowed, so that congruent shapes are also classed as similar. Non-uniform Scaling is obtained when at least one of the scaling factors is different from the other. Non-uniform scaling changes the shape of the object; e.g. a square may change into a rectangle or into a parallelogram if the sides of the square are not parallel to the scaling axes. For this specific reason we used Uniform Scaling to scale the figures in our 3-D model.

In our project Scaling is performed by the mere increasing or decreasing the area covered by the Red colour bands on our fingers. In each frame we obtained and stored the respective area of the figure using MATLAB functions and then we calculated the modified area of the same figure in the next frame. Then we obtained the difference between them and accordingly scaled the figure by updating the scaling parameter in the V-Realm Builder.

B. Translation: In Euclidean geometry, a Translation is a function that moves every point a constant distance in a specified direction. A translation can be described as a rigid motion: other rigid motions include rotations and reflections. A translation can also be interpreted as the addition of a constant vector to every point, or as shifting the origin of the coordinate system.

If we take the concept of translation and apply it to our 3-D models then we can simply infer that translation is nothing but the change in position of the centroid of a figure and the distance covered during the translation motion is nothing but the distance between the new centroid and the old centroid. Hence in our project for translation we simply tracked the centroid of each figure formed in their respective frames and updated the V-Realm Builder with the same value.
III. GESTURE BASED APPLICATIONS

Gesture based applications are broadly classified into two groups on the basis of their purpose: multi directional control and a symbolic language. A gesture recognition system could be used in any of the following areas:

A. 3D Design: CAD (computer aided design) is an HCI which provides a platform for interpretation and manipulation of 3-Dimensional inputs which can be the gestures. Manipulating 3D inputs with a mouse is a time consuming task as the task involves a complicated process of decomposing a six degree freedom task into at least three sequential two degree tasks. Massachusetts Institute of Technology has come up with the 3DRAW technology that uses a pen embedded in polhemus device to track the pen position and orientation in 3D. A 3space sensor is embedded in a flat palette, representing the plane in which the objects rest. The CAD model is moved synchronously with the users gesture movements and objects can thus be rotated and translated in order to view them from all sides as they are being created and altered.

B. Tele-presence: There may raise the need of manual operations in some cases such as system failure or emergency hostile conditions or inaccessible remote areas. Often it is impossible for human operators to be physically present near the machines. Tele presence is an area of technical intelligence which aims to provide physical operation support that maps the operator arm to the robotic arm to carry out the necessary task, for instance the real time ROBOGEST system constructed at University of California, San Diego presents a natural way of controlling an outdoor autonomous vehicle by use of a language of hand gestures. The prospects of Tele presence includes space, undersea mission, medicine manufacturing and in maintenance of nuclear power reactors.

C. Virtual reality: Virtual reality is applied to computer-simulated environments that can simulate physical presence in places in the real world, as well as in imaginary worlds. Most current virtual reality environments are primarily visual experiences, displayed either on a computer screen or through special stereoscopic displays. There are also some simulations include additional sensory information, such as sound through speakers or headphones. Some advanced, haptic systems now include tactile information, generally known as force feedback, in medical and gaming applications.

D. Sign Language: Sign languages are the most raw and natural form of languages could be dated back to as early as the advent of the human civilization, when the first theories of sign languages appeared in history. It has started even before the emergence of spoken languages. Since then the sign language has evolved and been adopted as an integral part of our day to day communication process. Now, sign languages are being used extensively in international sign use of deaf and dumb, in the world of sports, for religious practices and also at work places. Gestures are one of the first forms of communication when a child learns to express its need for food, warmth and comfort. It enhances the emphasis of spoken language and helps in expressing thoughts and feelings effectively. A simple gesture with one hand has the same meaning all over the world and means either ‘hi’ or ‘goodbye’. Many people travel to foreign countries without knowing the official language of the visited country and still manage to perform communication using gestures and sign language. These examples show that gestures can be considered international and used almost all over the world. In a number of jobs around the world gestures are means of communication.
E. Computer games: Using the hand to interact with computer games would be more natural for many applications. In our program we have made several computer games where a house has to be made by moving basic building box using hand gesture recognition.

![3D house building game screenshot](image)

**Fig.7** A screenshot of the 3D house building game

F. Mouse pointer control: Mouse pointer or any other pointer device movement can be controlled by recognize hand gesture and movement. In our project we have developed a simple painting software by which basic paint can be done only by recognising hand gesture, so we don’t need to use our mouse at all.

![Basic paint software screenshot](image)

**Fig.8** A screenshot of the basic paint software in which BITM SEF-2013 is written only by recognizing hand movement
IV. ADVANTAGES & DISADVANTAGES

The various advantages & disadvantages of Gesture Recognition are given below:

A. Advantages:
   - Speed and sufficient reliable for recognition system. Good performance system with complex background.
   - The radial form division and boundary histogram for each extracted region, overcome the chain shifting problem, and variant rotation problem.
   - Exact shape of the hand obtained led to good feature extraction. Fast and powerful results from the proposed algorithm.
   - Simple and active, and successfully can recognize a word and alphabet. Automatic sampling, and augmented filtering data improved the system performance.
   - The system successfully recognized static and dynamic gestures. Could be applied on a mobile robot control.
   - Simple, fast, and easy to implement. Can be applied on real system and play games.
   - No Training is required.

B. Disadvantages:
   - Irrelevant object might overlap with the hand. Wrong object extraction appeared if the objects larger than the hand. Performance recognition algorithm decreases when the distance is greater than 1.5 meters between the user and the camera.
   - System limitations restrict the applications such as; the arm must be vertical, the palm is facing the camera and the finger colour must be basic colour such as either red or green or blue.
   - Ambient light affects the colour detection threshold.
CONCLUSION

In today’s digitized world, processing speeds have increased dramatically, with computers being advanced to the levels where they can assist humans in complex tasks. Yet, input technologies seem to cause a major bottleneck in performing some of the tasks, under-utilizing the available resources and restricting the expressiveness of application use. Hand Gesture recognition comes to rescue here. Computer Vision methods for hand gesture interfaces must surpass current performance in terms of robustness and speed to achieve interactivity and usability. A review of vision-based hand gesture recognition methods has been presented. 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. The aim of this project was to develop an online Gesture recognition system and by doing that control various real time applications, like painting using our fingers and playing 3-D games without actually interacting with the computer. We have shown in this project that online gesture recognition system can be designed using Blob detection and tracking technique. The processing steps to classify a gesture included gesture acquisition, segmentation, morphological filtering, blob detection and classification using different technique. A threshold for Red colour is used for segmentation purpose and grey scale images is converted into binary image consisting hand or background. Morphological filtering techniques are used to remove noises from images so that we can get a smooth contour. The main advantage of Blob detection is that it is invariant to rotation, translation and scaling so it is a good feature to train the learning machine. We have achieved a good amount of accuracy using the above method. Area of Hand gesture based computer human interaction is very vast. This project recognizes hand gesture online so work can be done to do it for real time purpose. Hand recognition system can be useful in many fields like robotics, computer human interaction.

REFERENCES