International Journal of Scientific Research in Science and Technology
Print ISSN: 2395-6011 | Online ISSN: 2395-602X (www.ijsrst.com)
doi : https://doi.org/10.32628/IJSRST

Volume Controlling with Hand Gesture

Sonali Thanambir, Abhishek Shete, Trupti Mane, Harshada Gawade
U.G. Student, Department of Computer Engineering, Zeal College of Engineering & Research Pune, Maharashtra, India

Article Info
Volume 9, Issue 2
Page Number : 492-499

Publication Issue
March-April-2022

Article History
Accepted : 03 April 2022
Published : 20 April 2022

ABSTRACT

The project introduces an application using computer vision for Hand gesture recognition. A camera records a live video stream, from which a snapshot is taken with the help of interface. The system is trained for each type of count hand gestures (one, two, three, four, and five) at least once. After that a test gesture is given to it and the system tries to recognize it.

A research was carried out on a number of algorithms that could best differentiate a hand gesture. It was found that the diagonal sum algorithm gave the highest accuracy rate. In the preprocessing phase, a self-developed algorithm removes the background of each training gesture. After that the image is converted into a binary image and the sums of all diagonal elements of the picture are taken. This sum helps us in differentiating and classifying different hand gestures.

Previous systems have used data gloves or markers for input in the system. I have no such constraints for using the system. The user can give hand gestures in view of the camera naturally. A completely robust hand gesture recognition system is still under heavy research and development; the implemented system serves as an extendible foundation for future work.

KEYWORDS: Hand gesture, Recognition, Detection, Diagonal Sum, Pre-processing, Feature extraction, Skin Modeling, Labeling

I. INTRODUCTION

Recent developments in computer software and related hardware technology have provided a value added service to the users. In everyday life, physical gestures are a powerful means of communication. They can economically convey a rich set of facts and feelings. For example, waving one's hand from side to side can mean anything from a "happy goodbye" to "caution". Use of the full potential of physical gesture is also something that most human computer dialogues lack.

The task of hand gesture recognition is one the important and elemental problem in computer vision. With recent advances in information technology and media, automated human interactions systems are build which involve hand processing task like hand detection, hand recognition and hand tracking.

This prompted my interest so I planned to make a software system that could recognize human gestures through computer vision, which is a sub field of artificial intelligence. The purpose of my software through computer vision was to program a computer to "understand" a scene or features in an image.

A first step in any hand processing system is to detect and localize hand in an image. The hand detection task was however challenging because of variability in the pose, orientation, location and scale. Also different lighting conditions add further variability.

II. DIGITAL IMAGE PROCESSING

Image processing is reckoned as one of the most rapidly involving fields of the software industry with growing applications in all areas of work. It holds the possibility of developing the ultimate machines in future, which would be able to perform the visual function of living beings. As such, it forms the basis of all kinds of visual automation.

III. BIOMETRICS

Biometric systems are systems that recognize or verify human beings. Some of the most important biometric features are based physical features like hand, finger, face and eye. For instance finger print recognition utilizes of ridges and furrows on skin surface of the palm and fingertips. Hand gesture detection is related to the location of the presence of a hand in still image or in sequence of images i.e. moving images. Other biometric features are determined by human behavior like voice, signature and walk. The way humans generate sound for mouth, nasal cavities and lips is used for voice recognition. Signature recognition looks at the pattern, speed of the pen when writing ones signature.

IV. HAND GESTURE DETECTION AND RECOGNITION

DETECTION

Hand detection is related to the location of the presence of a hand in a still image or sequence of images i.e. moving images. In case of moving sequences it can be followed by tracking of the hand in the scene but this is more relevant to the applications such as sign language. The underlying concept of hand detection is that human eyes can detect objects which machines cannot with that much accuracy as that of a human. From a machine point of view it is just like a man fumble around with his senses to find an object.

The factors, which make the hand detection task difficult to solve, are:

Variations in image plane and pose

The hands in the image vary due to rotation, translation and scaling of the camera pose or the hand itself. The rotation can be both in and out of the plane.

Skin Color and Other Structure Components

The appearance of a hand is largely affected by skin color, size and also the presence or absence of additional features like hairs on the hand further adds to this variability.

Lighting Condition and Background

As shown in Figure 1.1 light source properties affect the appearance of the hand. Also the background, which defines the profile of the hand, is important and cannot be ignored.

Figure: 1.1: Lighting Condition and Background

RECOGNITION

Hand detection and recognition have been significant subjects in the field of computer vision and image processing during the past 30 years. There have been considerable achievements in these fields and numerous approaches have been proposed. However, the typical procedure of a fully automated hand gesture recognition system can be illustrated in the Figure 1.2 below:

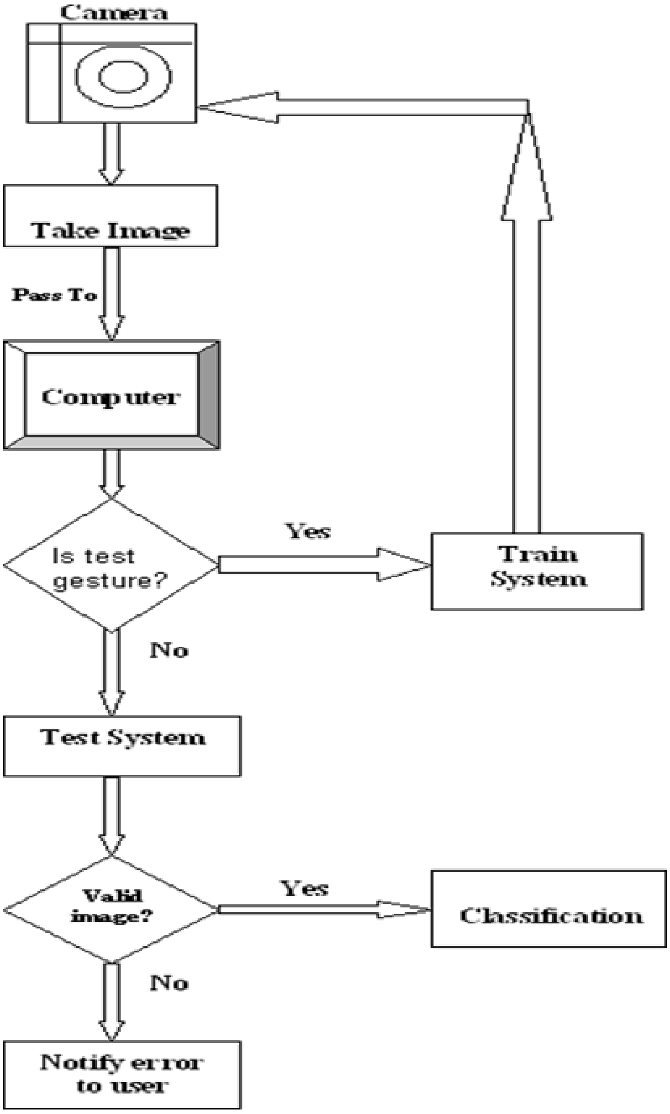

Figure 1.2: Hand Gesture Recognition Flow Chart

V. LITERATURE REVIEW

Hand gesture recognition research is classified in three categories. First **"Glove based Analysis"** attaching sensor with gloves mechanical or optical to transduces flexion of fingers into electrical signals for hand posture determination and additional sensor for position of the hand. This sensor is usually an acoustic or a magnetic that attached to the glove. Look-up table software toolkit provided for some applications to recognize hand posture.

The second approach is **"Vision based Analysis"** that human beings get information from their surroundings, and this is probably most difficult approach to employ in satisfactory way. Many different implementations have been tested so far.

One is to deploy 3-D model for the human hand. Several cameras attached to this model to determine parameters corresponding for matching images of the hand, palm orientation and joint angles to perform hand gesture classification. Lee and Kunii developed a hand gesture analysis system based on a three-dimensional hand skeleton model with 27 degrees of freedom. They incorporated five major constraints based on the human hand kinematics to reduce the model parameter space search. To simplify the model matching, specially marked gloves were used [3].

The Third implementation is **"Analysis of drawing gesture"** use stylus as an input device. These drawing analysis lead to recognition of written text. Mechanical sensing work has used for hand gesture recognition at vast level for direct and virtual environment manipulation. Mechanically sensing hand posture has many problems like electromagnetic noise, reliability and accuracy. By visual sensing gesture interaction can be made potentially practical but it is most difficult problem for machines.

LIGHTING

The task of differentiating the skin pixels from those of the background is made considerably easier by a careful choice of lighting. According to Ray Lockton, if the lighting is constant across the view of the camera then the effects of self-shadowing can be reduced to a minimum [25].

The top three images were lit by a single light source situated off to the left. A self- shadowing effect can be seen on all three, especially marked on the right image where the hand is angled away from the source. The bottom three images are more uniformly lit, with little self-shadowing. Cast shadows do not affect the skin for any of the images and therefore should not degrade detection. Note how an increase of illumination in the bottom three images results in a greater contrast between skin and background [25].

The intensity should also be set to provide sufficient light for the CCD in the camera. However, since this system is intended to be used by the consumer it would be a disadvantage if special lighting equipment were required. It was decided to attempt to extract the hand information using standard room lighting. This would permit the system to be used in a non-specialist environment [25].

BODY GESTURES

This section includes tracking full body motion, recognizing body gestures, and recognizing human activity. Activity may be defined over a much longer period of time than what is normally considered a gesture; for example, two people meeting in an open area, stopping to talk and then continuing on their way may be considered a recognizable activity. Bobick proposed taxonomy of motion understanding in terms of: Movement – the atomic elements of motion, Activity – a sequence of movements or static configurations and Action – high-level description of what is happening in context.

Most research to date has focused on the first two levels [3].

- The Pfinder system developed at the MIT Media Lab has been used by a number of groups to do body tracking and gesture recognition. It forms a two-dimensional representation of the body, using statistical models of color and shape. The body model provides an effective interface for applications such as video games, interpretive dance, navigation, and interaction with virtual characters [3].
- Lucente combined Pfinder with speech recognition in an interactive environment called Visualization Space, allowing a user to manipulate virtual objects and navigate through virtual worlds [3].

VISION-BASED GESTURE RECOGNITION

The most significant disadvantage of the tracker-based systems is that they are cumbersome. This detracts from the immerse nature of a virtual environment by requiring the user to put on an unnatural device that cannot easily be ignored, and which often requires significant effort to put on and calibrate. Even optical systems with markers applied to the body suffer from these shortcomings, albeit not as severely. What many have wished for is a technology that provides real- time data useful for analyzing and recognizing human motion that is passive and non-obtrusive. Computer vision techniques have the potential to meet these requirements.

Vision-based interfaces use one or more cameras to capture images, at a frame rate of 30 Hz or more, and interpret those images to produce visual features that can be used to interpret human activity and recognize gestures [3].

Typically the camera locations are fixed in the environment, although they may also be mounted on moving platforms or on other people. For the past decade, there has been a significant amount of research in the computer vision community on detecting and recognizing faces, analyzing facial expression, extracting lip and facial motion to aid speech recognition, interpreting human activity, and recognizing particular gestures [3].

Currently, most computer vision systems use cameras for recognition. Analog cameras feed their signal into a digitizer board, or frame grabber, which may do a DMA transfer directly to host memory. Digital cameras bypass the analog-to-digital conversion and go straight to memory. There may be a preprocessing step, where images are normalized, enhanced, or transformed in some manner, and then a feature extraction step. The features

which may be any of a variety of two- or three-dimensional features, statistical properties, or estimated body parameters – are analyzed and classified as a particular gesture if appropriate [3].

This technique was also used by us for recognizing hand gestures in real time. With the help of a web camera, I took pictures of hand on a prescribed background and then applied the classification algorithm for recognition.

PROJECT CONSTRAINTS

I propose a vision-based approach to accomplish the task of hand gesture detection. As discussed above, the task of hand gesture recognition with any machine learning technique suffers from the variability problem. To reduce the variability in hand recognition task we assume the following assumptions:

- Single colored camera mounted above a neutral colored desk.
- User will interact by gesturing in the view of the camera.
- Training is must.
- Hand will not be rotated while image is capturing.

The real time gesture classification system depends on the hardware and software.

Hardware

- Minimum 2.8 GHz processor Computer System or latest
- 52X CD-ROM drive
- Web cam (For real-time hand Detection)

Software

- Windows 2000(Service Pack 4),XP, Vista or Windows 7
- Matlab 8.0 or latest (installed with image processing toolbox)
- Vcapg2.dll (Video Capture Program Generation 2)
- DirectX 9.0 (for supporting Vcapg2)

VI. BRIEF OUTLINE OF THE IMPLEMENTED SYSTEM

Hand gesture recognition system can be divided into following modules:

- Preprocessing
- Feature extraction of the processed image
- Real time classification

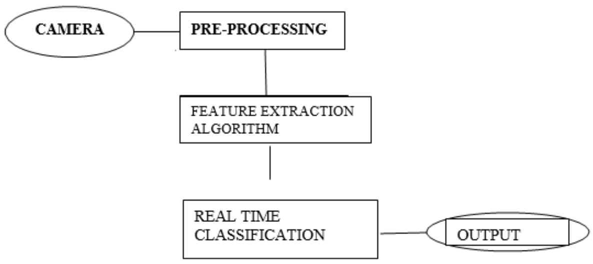

FIG. System Implementation:

FEATURE EXTRACTION ALGORITHMS

There are four types of algorithms that I studied and implemented namely as followings:

- **Row vector algorithm**
- **Edging and row vector passing**
- **Mean and standard deviation of edged image**
- **Diagonal sum algorithm**

ROW VECTOR ALGORITHM

We know that behind every image is a matrix of numbers with which we do manipulations to derive some conclusion in computer vision. For example we can calculate a row vector of the matrix. A row vector is basically a single row of numbers with resolution 1*Y, where Y is the total no of columns in the image matrix. Each element in the row vector represents the sum of its respective column entries as illustrated in Figure 4.4:

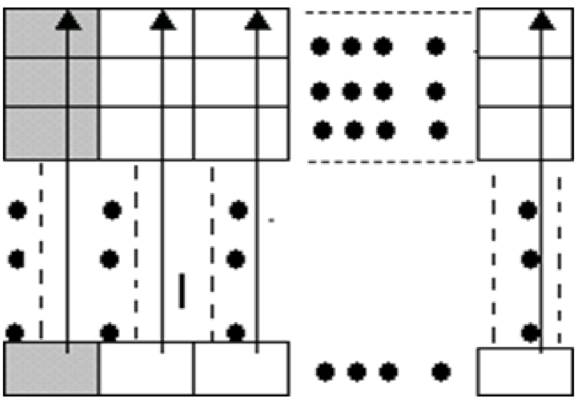

Figure 4.4 Row vector of an image

EDGING AND ROW VECTOR PASSING ALGORITHM

In the pre-processing phase of this algorithm, I do preprocessing, skin modeling and removed the background etc. of the gesture image taken. This image was converted from RGB into grayscale type. Gray scale images represent an image as a matrix where every element has a value corresponding to how bright/dark the pixel at the corresponding position should be colored.

MEAN AND STANDARD DEVIATION OF EDGED IMAGE

In the pre-processing phase, doing several step like removing the background and RGB image is converted into grayscale type as done in the previous algorithm. The edge of the grayscale image is taken with a fixed threshold i.e. 0.5 then calculate the mean and standard deviation the processed image.

Mean is calculated by taking a sum of all the pixel values and dividing it by the total no of values in the matrix. Mathematically, it is defined as:

$$\overline{X} = \sum_{i=1}^{n} Xi / n$$

Stand Deviation can calculate from mean which is mathematically defined as:

$$s = \sqrt{\frac{\sum_{i=1}^{n} (x_i - \overline{x})^2}{n}}$$

DIAGONAL SUM ALGORITHM

In the pre-processing phase, doing mentioned steps in methodology, skin modeling removal of the background, conversion of RGB to binary and labeling. The binary image format also stores an image as a matrix but can only color a pixel black or white (and nothing in between). It assigns a 0 for black and a 1 for white. In the next step, the sum of all the elements in every diagonal is calculated. The main diagonal is represented as k=0 in Figure 4.7 given below; the diagonals below the main diagonal are represented by k<0 and those above it are represented as k>0

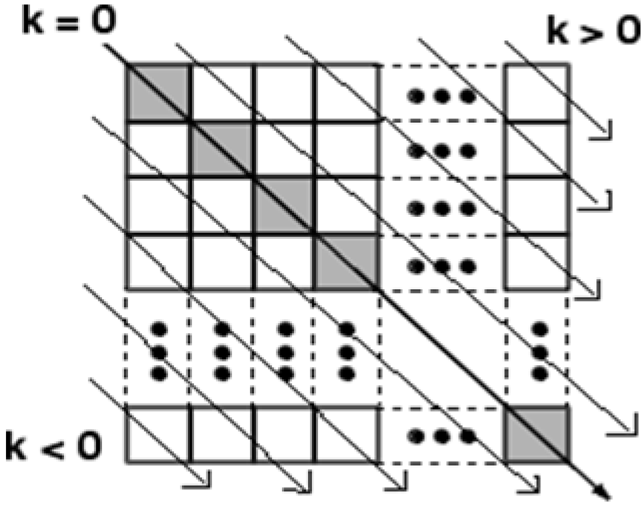

Figure: 4.8: Diagonal Sum

VII. CONCLUSION AND FUTURE WORK

This chapter summarizes my work at every stage of the project. At the time I started my thesis, I had a brief idea of how I will bring it from a topic on the paper to a real product. Due to knowledge of Computer Vision and Biometric subjects I had background in the image-processing field but not at expert level but my constant effort helped me to go through and succeed eventually.

As required in every project, research is of utmost importance. So, I spent the pretty much time in going through the background literature. I looked at various approaches of doing my thesis and developed four different methods: Row vector algorithm, Edging and row vector passing algorithm, Mean and standard deviation of edged image and Diagonal sum algorithm. Each of these algorithms was tried with neural networks and have higher performance rate in the ascending order respectively.

The first limitation that was discovered in all the algorithms used with neural networks was that their performance depended on the amount of training dataset provided. The system worked efficiently after being trained by a larger dataset as compared to a smaller dataset.

Due to the unusual behavior of neural network with all the mentioned parameters, the diagonal sum algorithm was finally implemented in real time. The system was tested with 60 pictures and the detection rate was found to be 86%. The strengths and weaknesses of gesture recognition using diagonal sum have been presented and discussed. With the implemented system serving as an extendible foundation for future research, extensions to the current system have been proposed.

VIII. FUTURE WORK

The system could also be made smart to be trained for only one or two gestures rather than all and then made ready for testing. This will require only a few changes in the current interface code, which were not performed due to the shortage of time.

One time training constraint for real time system can be removed if the algorithm is made efficient to work with all skin types and light conditions which seems impossible by now altogether. Framing with COG (Centre of gravity) to control orientation factor could make this system more perfect for real application.

IX. REFERENCES

[1]. J.Jenkinwinston (96207106036), M.Maria Gnanam (96207106056), R.Ramasamy (96207106306), Anna University of Technology, Tirunelveli: Hand Gesture Recognitionsystem Using Haar Wavelet.

[2]. Kay M. Stanney HANDBOOK OF VIRTUAL ENVIRONMENTS Design, Implementation, and Applications, Gesture Recognition Chapter #10 by Matthew Turk.

[3]. Daniel Thalman, Gesture Recognition Motion Capture, Motion Retargeting, and Action Recognition

[4]. Prof. Aparna Mote , Jagdish Patil ,Gajanan Sainde ,Shailendra Menge,"Thumb Based Document Retrieval Using Cloud System", International Journal of General Science and Engineering Research (IJGSER), ISSN 2455-510X, Vol 4(1), 2018, 27-2

[5]. J. Heinzm ann and A. Zelinsky Robust Real - Time Face Tracking and Gesture Recognition

[6]. Vladimir Vezhnevets, Vassili Sazonov, Alla Andreeva, Moscow State University A Survey on Pixel-Based Skin Color Detection Techniq.

[7]. S. Barber et al, "Bitter to better-how to make bitcoin a better currency," in Proc. Int. Conf. Financial Cryptography Data Security, 2012, pp. 399–414.

[8]. Ahuja M.K., Singh A. Static vision based Hand Gesture recognition using principal component analysis; Proceedings of the 2015 IEEE 3rd International Conference on MOOCs, Innovation and Technology in Education (MITE); Amritsar, India. 1–2 October 2015; pp. 402–406. [Google Scholar

[9]. Dave Bryson Dave Penny David C. Goldenberg Gloria Serrao, "BLOCKCHAIN TECHNOLOGY FOR GOVERNMENT", MITRE December 2017.

[10]. Mitra S., Acharya T. Gesture recognition: A survey. IEEE Trans. Syst. Man Cybern. Part C Appl. Rev. 2007;37:311–324. doi: 10.1109/TSMCC.2007.893280. [CrossRef] [Google Scholar

[11]. I. Alqassem et al., "Towards reference architecture for cryptocurrencies: Bitcoin architectural analysis," in Proc. IEEE Internet Things, IEEE Int. Conf. Green Comput. Commun. IEEE Cyber, Physical Social Comput. 2014, pp. 436–443