Gesture Controlled Processing

A Report Submitted
in Partial Fulfillment of the Requirements
for the Degree of
Bachelor of Technology
in
Information Technology
by

Yash Rathore (20168086)
Siddhant (20168059)
Sarvottam (20168043)
Vignesh Rao (20168063)

under the guidance of
Er. Rajesh Tripathi
(Associate Professor)

to the

COMPUTER SCIENCE AND ENGINEERING DEPARTMENT
MOTILAL NEHRU NATIONAL INSTITUTE OF TECHNOLOGY, ALLAHABAD
PRAYAGRAJ, UTTAR PRADESH (INDIA)
April 2019

UNDERTAKING

I declare that the work presented in this report titled "*Gesture Controlled Processing*", submitted to the Computer Science and Engineering Department, Motilal Nehru National Institute of Technology, Allahabad, Prayagraj, Uttar Pradesh (India) for the award of the ***Bachelor of Technology*** degree in ***Information Technology***, is my original work. I have not plagiarized or submitted the same work for the award of any other degree. In case this undertaking is found incorrect, I accept that my degree may be unconditionally withdrawn.

April 2019
Prayagraj
U.P. (India)

(Yash Rathore, 20168086)

(Siddhant, 20168059)

(Sarvottam, 20168043)

(Vignesh Rao, 20168063)

'

CERTIFICATE

Certified that the work contained in the report titled "*Gesture Controlled Processing*", by *Yash Rathore (20168086), Siddhant (20168043), Sarvottam (20154095), Vignesh Rao (20168063)*, has been carried out under my supervision and that this work has not been submitted elsewhere for a degree.

Er. Rajesh Tripathi
(Associate Professor)
Computer Science and Engineering Dept.
M.N.N.I.T, Allahabad

April 2019 — Prayagraj, Uttar Pradesh (INDIA)

Preface

Gesture recognition is a topic which aims to interpret human gestures via mathematical algorithms. One of the main focuses in the field is emotion recognition through face and hand gesture recognition. One of the main goals today is to overcome the challenge of making interactions more easy, natural, and convenient without wearing any extra device. Having the potential for its application in user interfaces, recent years have seen major growth in the technology. Growth in innovation and technology has given a mammoth impetus to automation in every field -dependence on human control sequence is reducing in order to do away with the drawbacks of manual control. Keeping this in mind, our focus is on creating a system wherein human gestures can control various processes for instance, gesture-controlled/ automated rooms, keyboards, mouse et al. It basically involves study and application of various concepts of Image Processing for the achievement of the aforesaid goal. Image Processing here has been applied for implementation of Virtual mouse and Virtual keyboard which will be the next step in the evolution of the compact electronics. The vision of this project was to create a real time system which can be used for uniquely identifying various gestures and performing multiple task using that. The project also aims to use only the open source tool so that the cost associated with the product remains as low as possible.

Acknowledgements

The completion of this project required a lot of effort, guidance and support from many people. We feel privileged and honoured to have got this all along the development of the project.

We would like to express our gratitude to our mentor, Er. Rajesh Tripathi (Associate Professor) for his perennial support, guidance and monitoring throughout the making of this project. Also, we would also like to thank Dr Shashank Srivastava, DUGC Convener for his exemplary support and motivation for bringing about the possibility of projects for CSE Department. We would also like to acknowledge and thank our professors, colleagues and seniors for supporting us and enabling us to opt a different approach to our project.

Contents

Chapter 1

Introduction

Gesture recognition is a topic which aims to interpret human gestures via mathematical algorithms. Even though gestures can refer to any kind of bodily motion, they commonly originate from the face or hand. Currently, the field is focused on emotion recognition from face and hand gesture recognition. Both engineers and scientists are equally involved in the field that is gesture recognition. It is the natural way of human machine interaction. One of the main goals today is to overcome the challenge of making interactions more easy, natural, and convenient without wearing any extra device. It is of great need in designing an efficient human-computer interface. Having the potential for its application in user interfaces, recent years have seen major growth in the technology.

In gesture recognition, the gestures can be transmitted using special gloves, or they can also be read using a camera. The captured data is processed and used as an input to handle applications or devices. The project entails the use of image processing for gesture recognition and utilizing the result of it for multilateral purposes. This project is a real time system where image processing and machine learning is used for recognizing the face and hand gesture and using the recognized gesture for performing multiple tasks like virtual keyboard control and virtual mouse control.

1.1 Motivation

Our focus is on creating a system wherein human gestures can control various processes for instance, gesture-controlled/automated rooms, keyboards, mouse et al. Such systems shall also aim to simplify life of physically impaired individuals by helping them in their daily chores. Also there could be a situation where the hardware is already in place but the platform on which the gesture recognition software is built on, doesn't support the existing hardware, which would lead to extra hardware cost. To tackle this problem, the existing solutions come with a pre-defined list of hardware with which they are compatible. Despite considerable advances in recent years, there are still challenges in full body gesture recognition behaviour understanding. Our work is fully focused on hand gesture recognition. The challenges for gesture recognition are position of the individual relative to the cameras, illumination changes, noisy videos, robustness, partial or full occlusion, handling a wide range of human pose configurations, huge variability of the 2D appearance of unexpected objects, different points of view, etc. Some of these challenges can be handled by designing a video based gesture recognition system that is highly accurate, fast, secure and easily scalable to a large number of postures. The need is to focus on the specific application and develop the gesture recognition algorithm

1.2 Basic Concepts

1.2.1 Gesture

A gesture is a form of non-verbal communication or non-vocal communication in which visible bodily actions communicate particular messages, either in place of, or in conjunction with, speech. Gestures may include any kind of bodily actions, ranging from but not limited to movement of hands, facial emotions, etc. Gestures allow a person to communicate their feelings and thoughts, often together in addition to words when they speak.

1.2.2 Gesture Recognition

Gesture recognition is a process to recognize human gestures and process them so that the data can be used to perform some action or control a device. In gesture recognition, the gestures made by the hand (or other body parts) are read by a camera or transmitted using special hardware like a robotic glove. On the basis of data acquisition, the gesture recognition can be classified into two categories viz. Glove-based, and Vision based gesture recognition system. On the one hand, Glove-Based Gesture recognition uses specialized gloves or sensors for data processing and power supply. The glove worn by users, extract the configuration of the hand in the 3 dimensions along with the way it moves in space. The setting up of such systems takes a lot of time and thus are unfit for spontaneous interaction due to such engagement. For Example, Acceleglove created by George Washington University translates American Sign Language gestures into text. There are a series of accelerometers on each finger of the glove and other sensors on the shoulders and elbows to send electrical signals to a micro-controller. It then interprets the action associated with the gesture. On the other hand, Vision-based Gesture Recognition Systems use hardware equipments like the camera to extract the features of the hand/face depending on hand-gesture or facial emotion recognition. These however, have certain problems of their own, such as, different aspects of the environment ranging from the light conditions to the amount of (visual)noise, video camera settings, quality of the image being captured by the camera, etc. For example, in Kinect for Windows applications, the user can interact naturally with the computer using simple gestures. Xbox Project Natal allows the user to interact with the display without requiring any controller which gives a feeling of natural interaction. To kick a ball on the screen, the user will have to perform a kick motion.

1.3 Methodology

1. Image Processing: The extracted snaps will go through image processing for detecting the gesture. First we do background subtraction, then the

segmented hand region is processed to find the gesture

2. Virtual Keyboard/Mouse Control: Various combination of gestures will be used for virtually controlling keyboard and mouse.

Chapter 2

Image Processing

2.1 Image

An image is defined as a two-dimensional function,F(x,y), where x and y are spatial coordinates, and the amplitude of F at any pair of coordinates (x,y) is called the intensity of that image at that point. In other words, an image can be defined by a two-dimensional array specifically arranged in rows and columns. Digital image is a collection of a finite number of elements such that each element has a specified location and a specified value, changing which would change the whole image.These elements are referred to as picture elements,image elements,and pixels. A pixel can be defined as the smallest independent element of an image.

2.2 The HSV Color space

The HSV color space provides more intuition on how people experience color than the RGB color space does. As hue (H) varies from 0 to 1.0, the corresponding colors vary from red, through yellow, green, cyan, blue, and magenta, back to red. As saturation(S) varies from 0 to 1, the corresponding hue values change from unsaturated (visible as different shades of gray) to fully saturated hues(which contain no white component). As value(V), or brightness, varies from 0 to 1, the corresponding colors become increasingly brighter. The hue component in HSV is represented in a hexagon with the range 0 to 360 angle all lying around in it. With RGB the color will have values like (0.5, 0.5, 0.25), whereas for HSV it will be (30, 3/4, 0.5). HSV is best used when a user is selecting a color interactively It is usually much easier for a user to get to a desired color as compared to using RGB.

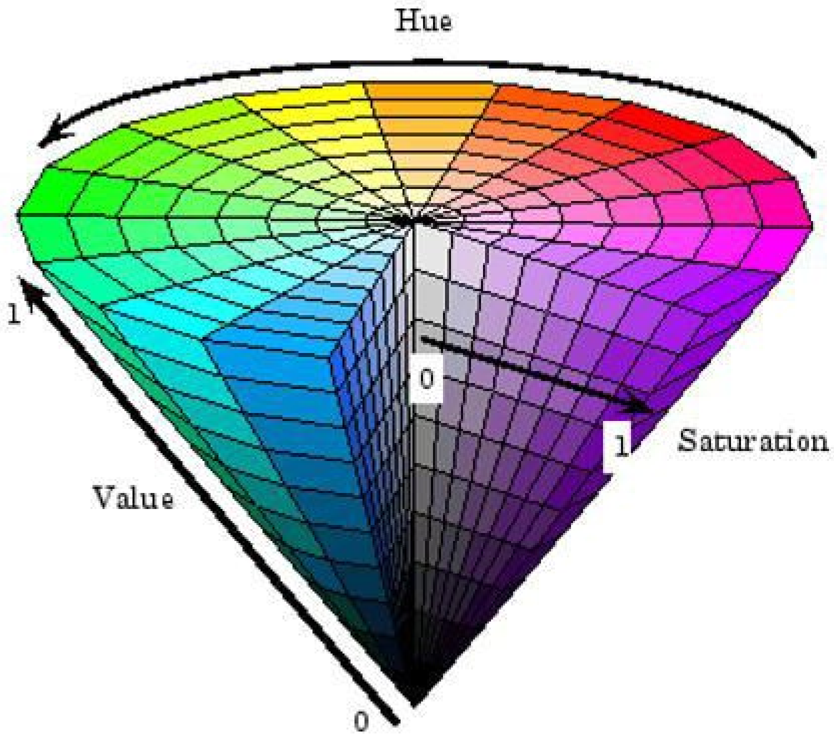


Figure 1: HSV Color Space

2.3 Image Pre-processing

Image preprocessing refers to operations on images at the lowest abstraction level of intensity of input and output images, to restrain errors related to brightness and geometry values of the pixels. Appropriate mathematical and statistical models(which can be definite or indefinite) are used to reduce such errors. To improve the visual impact of pixel intensity values, we use image enhancement, which is the modification of image by changing pixel brightness values. Image enhancement involves many techniques to improve the appearance of an image, or maybe to convert the image to a form which is better suited for interpretation by humans or machines. Images obtained from satellites and conventional and digital cameras sometimes lack in brightness and contrast because of the limitations of imaging sub systems and illumination conditions while capturing image. Images may contain different kinds of noise. In image enhancement is to accentuate certain image features for image display or subsequent analysis. Some examples are contrast and edge enhancement, sharpening, noise filtering, pseudo-coloring, and magnifying. Image enhancement is useful in feature extraction, image analysis and display. The enhancement process itself does not increase the inherent information content in the data. It simply emphasizes certain specified image characteristics. Enhancement algorithms are generally interactive and dependent on the application being used. Some of the enhancement techniques are:

2.3.1 Contrast Stretching

Some images like images containing water bodies, desert regions, dense forested areas, tundra or snowy regions, clouds and hazy conditions over heterogeneous regions are called as homogeneous i.e., they do not have much change in their levels. They are characterized by the very narrow peaks in the histogram representation. Incorrect illumination of the scene can also cause homogeneity of the image. Ultimately the images hence obtained are not easily interpretable due to poor human perceptibility. It is so because there exist only a narrow range of gray levels in the image having provision for wider range of gray-levels. The contrast stretching methods are designed exclusively for frequently encountered situations. Many stretching techniques have been developed to stretch the narrow range to a large available dynamic range.

2.3.2 Noise Filtering

Noise Filtering is used to filter any redundant or unnecessary information from an image which is referred to as noise. This feature is mostly interactive. Various filters are used like low pass filters, high pass filters, mean filters, median filters etc.

2.3.3 Histogram Modification

Histogram is of prime importance in image enhancement techniques. It reflects the characteristics of an image. By modifying the histogram, the respective image characteristics are modified. One such example is Histogram Equalization, which is used to regulate image contrast. It is a nonlinear stretch that rearranges pixel values such that there are approximately the equal number of pixels of each value within a range. This operation results in a flat histogram. Therefore, contrast is increased at the peaks and lessened at the tails.

2.4 Thresholding

Thresholding is one of the simplest methods of image segmentation. It is used to create binary images from gray scale images. A threshold constant T is fixed at the start of operation. The simplest thresholding method replaces each pixel in an image with a black pixel if the image intensity is less than this constant or a white pixel if the image intensity is greater than the constant. In the example image on the right, this results in the background of the hand image becoming completely black, and the greyish hand outline becoming completely white.

2.5 Skin Detection

Skin detection is the technique of highlighting skin-colored pixels in an image or a video. Skin detection is used in preprocessing to filter out regions potentially containing human faces, limbs or body in images. It is used in a wide range of image processing applications like skin disease detection, face recognition, gesture tracking and human-computer interaction, the latter two of which are the focus of this project. The primary feature for recognition of skin from an image can be the skin color. But skin color cannot be the only deciding feature due to the variation in skin tone according to different races. Other factors such as the lighting conditions also affect the results, and must be taken into consideration. Respecting these conditions, skin color is often combined with other cues like edge features and texture. This is attained by breaking down the image into individual pixels and categorizing them into skin colored or non-skin colored pixels. One simple method is to check if each skin colored pixel falls into a pre-defined color range or a region of values in some coordinates of a color space.

2.6 Convex Hull

A set of points is defined to be convex if it contains the line segments connecting each pair of its points. The convex hull of a given set X may be defined as one of the following four

1. It can be the unique minimal convex set containing X.
2. It can be the intersection of all convex sets containing X.
3. It can be the set of all convex combinations of points in X.
4. Or, it can be the union of all simplices with vertices in X.

It is unsure that the first definition makes much sense: why should there exist a unique minimal convex set containing X, for every X? However, the second definition, the intersection of all convex sets containing X, is more promising, defined better as a subset of every other convex set Y that contains X, because X is included among the sets being intersected. Thus, it is the unique minimal convex set containing X which is minimum in size. Each convex set containing X must (by the assumption that it is convex) contain all convex combinations of points in X, so the set of all convex combinations is contained in the intersection of all convex sets containing X. Conversely, the set of all convex combinations is itself a convex set containing X, so it also contains the intersection of all convex sets containing X, and therefore the sets given by these two definitions must be equal.

In fact, according to Carathodory's theorem, if X is a subset of an N- dimensional vector space, convex combinations of at most N + 1 points are sufficient in the definition above. Therefore, the convex hull of a set X of three or more points in the plane is the union of all the triangles determined by triples of points from X, and more generally in N-dimensional space the convex hull is the union of the simplices determined by at most N + 1 vertices from X. If the convex hull of X is a closed set (as happens, for instance, if X is a finite set or more generally a compact set), then it is the intersection of all closed half-spaces containing X. The hyperplane separation theorem proves that in this case, each point not in the convex hull can be separated from the convex hull by a half-space. However, there exist convex sets, and convex hulls of sets, that cannot be represented in this way. Open half spaces are such examples.

2.7 Contour Tracing

Contour tracing is a technique that is applied to images to extract the object boundaries. It is also known as border following. Once the contour area of a given pattern is extracted, it's different characteristics are examined and utilized as features which later are used in pattern classification and recognition. Therefore, proper extraction of contour area helps in producing better and more accurate features which increase the chances of correctly classifying patterns. The contour area is generally a small subset of the collective pixels representing a pattern. Therefore, the amount of computation is greatly reduced when we run feature extracting algorithms on the contour instead of on the whole pattern. Since the contour shares a lot of features with the original pattern, the feature extraction process becomes much more efficient when performed on the contour rather on the original pattern. In conclusion, contour tracing is almost always a major contributing procedure to the efficiency of the feature extraction process - an essential process in the field of pattern recognition.

2.8 Masking

Masking is the process of setting up of some pixel values in an image to zero, or some other "background" value. Masking can be done in two ways:

- By using an image as a mask. A mask image is an image where some of the pixel intensity values are zero, and others are non-zero. Wherever the pixel intensity is zero in the mask image, the pixel intensity of the resulting masked image will be set to the background value (normally zero).
- Using a set of region of interest as the mask. The region of interest for each part are used to define the mask.

2.9 Background Subtraction

Background subtraction is process of extracting foreground objects from maintained background model. Any entity that is detected by producing difference of the every frame of sequence to background model is called a foreground entity. The obtained result can be further used for tracking targets, motion detection. Background subtraction can be sub-classified into parametric and non-parametric background subtraction. There are several background subtractions techniques that are available for use. The background model can be static or dynamic. Figure 4 shows the flowchart for Background subtraction. The model in which the background of scene may contain moving objects in outdoor environment is called dynamic background model. Pixel- based and block based are two major kind of approached are for background Subtraction. Non- parametric statistical Modeling of pixel process is used to construct a statistical representation of background scene. Several challenges that have to be faced to construct a good background subtraction algorithm are robustness against the changes in illumination and shadow detection. Background subtraction is used for detecting moving objects in videos from static cameras. The method in the approach is that of detecting the moving objects from the difference between the current frame and a reference frame, i.e called the background image, or background model. The background image must be a representation of the scene with non-moving objects and must be regularly updated in order to adapt to the varying luminaries conditions and geometry settings. More complex models have extended the concept of background subtraction beyond its formal meaning. Various new methods have been proposed for performing background subtraction. All of these methods try to effectively estimate the background model from the temporary sequence of the frames. However, there is a wide variety of techniques and both the expert and the newcomer to this area can be confused about the benefits and limitations of each method. Below are few main methods used and their categorization.

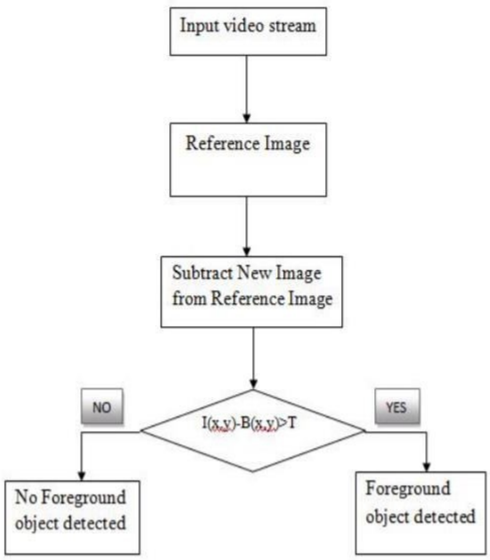


Figure 2: Background Subtraction

2.9.1 Temporal Average Filter

The temporal average filter is a method that was proposed at the Velastin. The background model is estimated from the median of all pixels of a number of previous images by this filter. In this system a buffer with the pixel values of the last frames is used to update the median for each image. The system examines all images in a given time period called training time to model the background. At this time we only display images and will find the median, pixel by pixel, of all the plots in the background this time. Each pixel value is compared with the input value of funds previously calculated, after the training period for each new frame. If the input pixel is within a required value, the pixel is considered to match the background model and its value is added in the pixbuf. Else the pixel is classified as foreground, and not included in the buffer. This method can not be considered very efficient because they do not present a rigorous statistical basis and requires a buffer that has a high computational cost.

2.10 Applications

- Color processing
- Video processing
- Pattern recognition
- Microscopic Imaging
- Others
- Remote sensing
- Image sharpening and restoration
- Machine/Robot vision
- Medical field
- Transmission and encoding

Chapter 3

Overview Of Project

3.1 Workflow

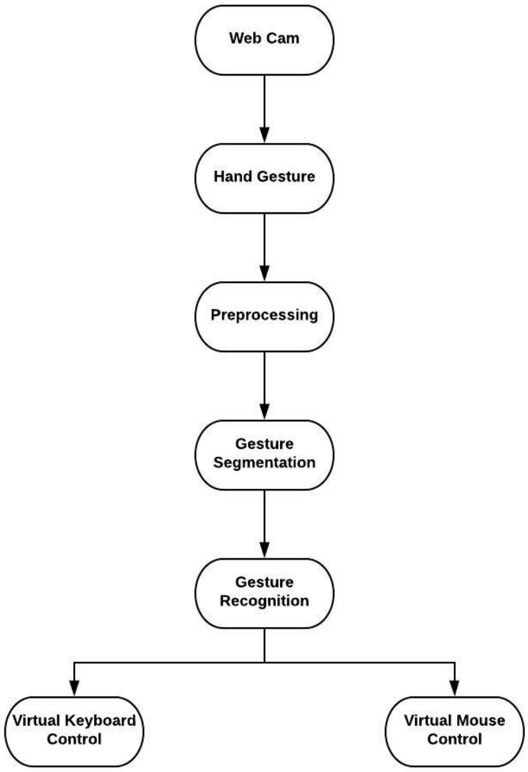


Figure 3: Workflow

3.2 Hand Gesture Recognition Process

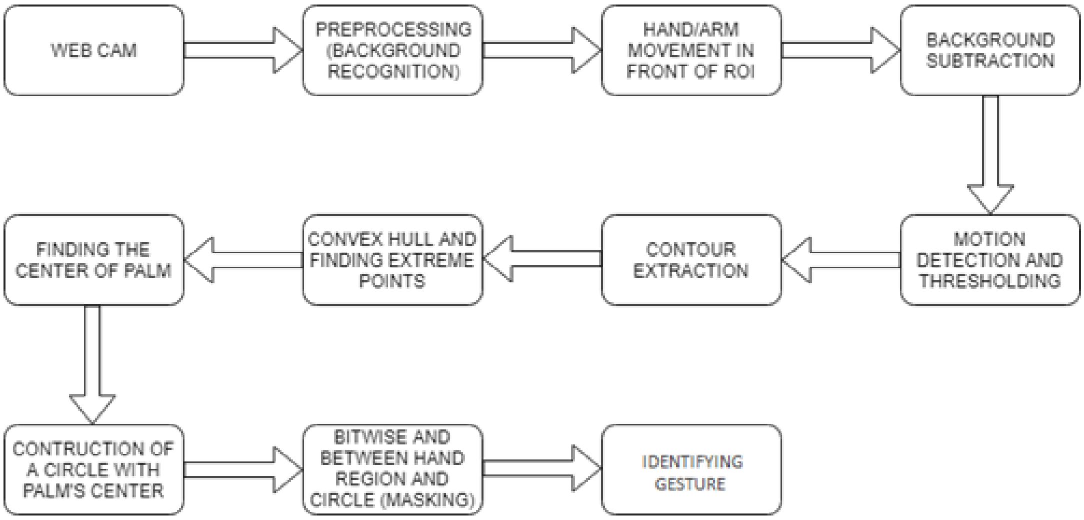


Figure 4: Hand Gesture Recognition Process

3.2.1 Preprocessing and Background Subtraction

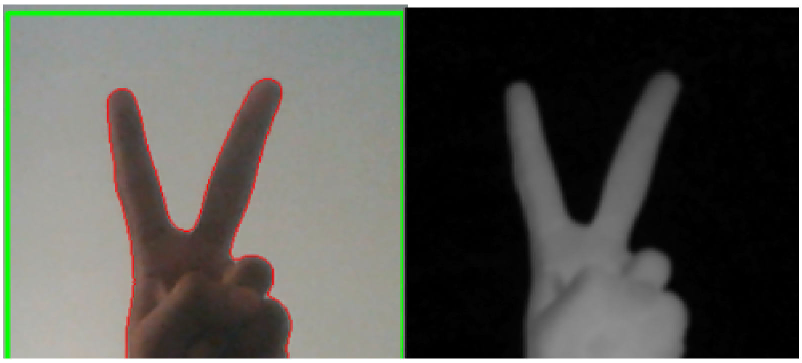


Figure 5: Background Subtraction

- We make our system to look over a particular scene for 30 frames.
- During this period, we compute the running average over the current frame and the previous frames
- After figuring out the background model using running averages, we use the current frame which holds the foreground object (hand in our case) in addition to the background.
- We calculate the absolute difference between the background model (updated over time) and the current frame (which has our hand) to obtain a difference image that holds the newly added foreground object (which is our hand).

3.2.2 Thresholding and Contour Extraction

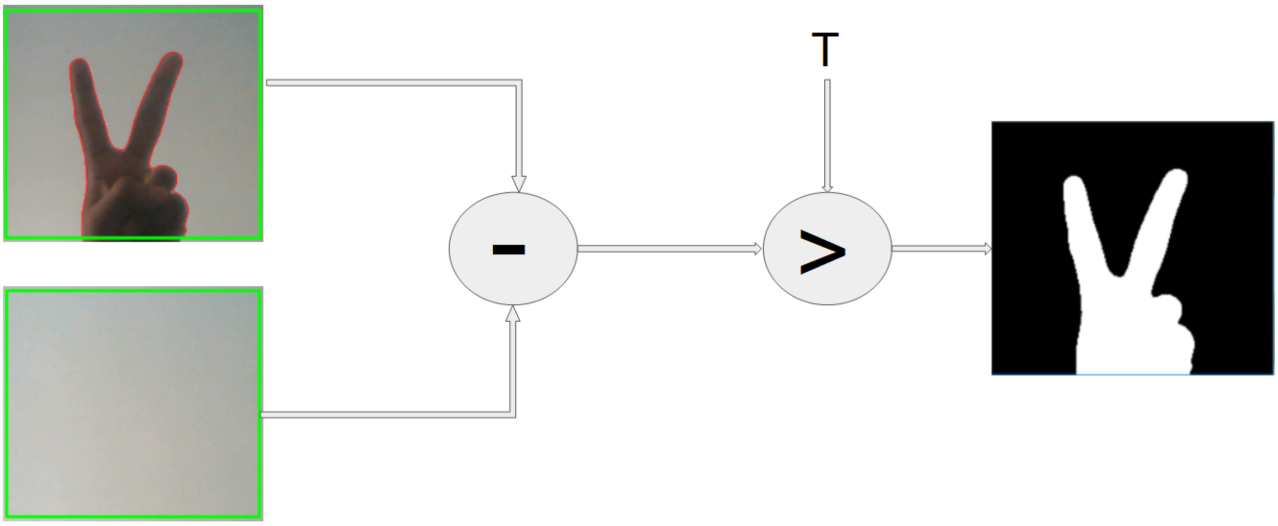


Figure 6: Thresholding

- After thresholding the difference image, we find contours in the resulting image. The contour with the largest area is assumed to be our hand.

3.2.3 Convex Hull, Finding the palm's centre, and construction of circle with assumed radius

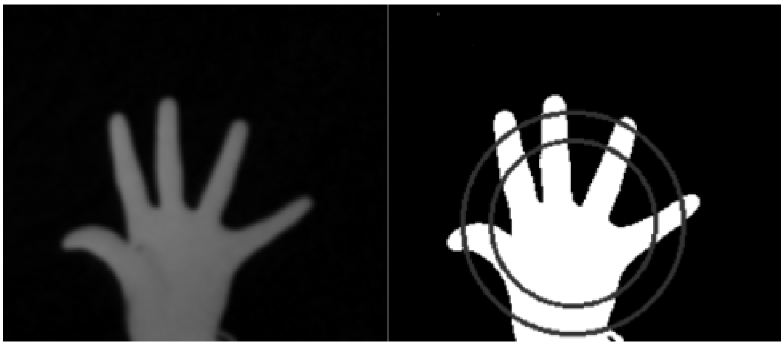


Figure 7: Constructing Circle

- Convex hull of the segmented hand region (which is a contour) is found and the most extreme points in the convex hull (Extreme Top, Extreme Bottom, Extreme Left, Extreme Right) are computed.
- Center of palm is found using these extremes points in the convex hull.
- Using the palms center, two circles are constructed with radius less than the maximum Euclidean distance (between the palms center and the extreme points).

3.2.4 Masking and Identifying the Gesture

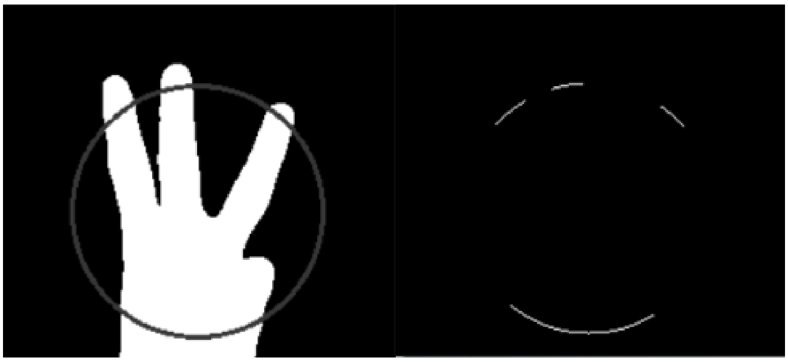


Figure 8: Masking and identifying Gesture

- Bitwise AND operation is performs between the thresholded hand image (frame) and the circular ROI (mask).
- Using the result various gestures can be recognized.

3.3 System Requirements

- Python 3.5.2 or higher
- Opencv 3.2.0
- Numpy 1.12.1
- Statistics
- Imutils

Chapter 4

Working And Results

4.1 Virtual Keyboard Control

4.1.1 Media Player Control

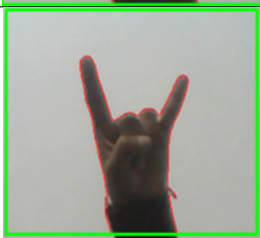
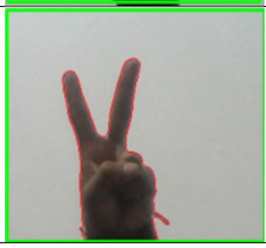
Gesture	Action Performed
	Video Starts
	Video Plays
	Video Pauses

Table 1: Mapping for Media Player Control

4.1.2 Alphabet Keypad Control

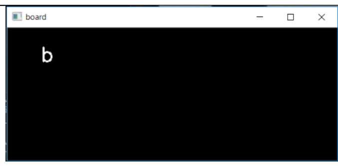
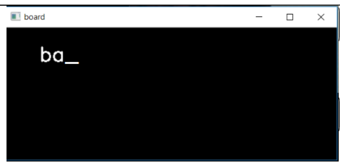
Gesture	Action
	
	
	
	
	
	

Table 2: Example of Alphabet Keypad Control

4.1.3 Calculator Control

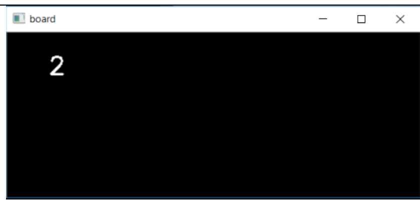
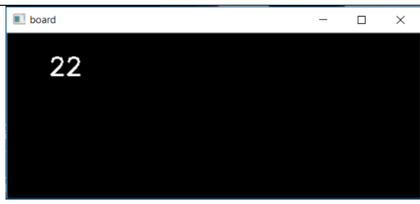
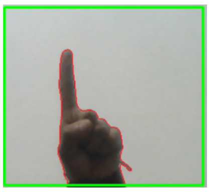
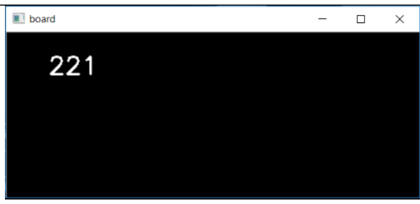
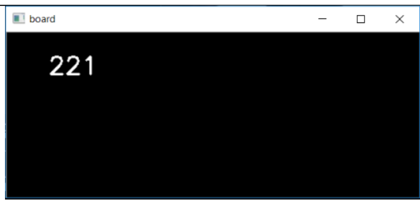
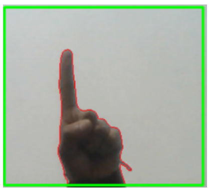
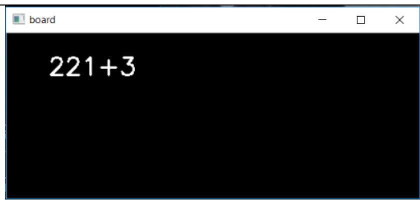
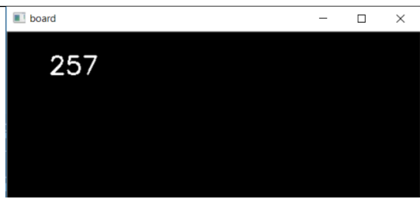
Gesture	Action
	
	
	
	
	
	
	
	

Table 3: Example of Calculator Control

4.2 Virtual Mouse Control

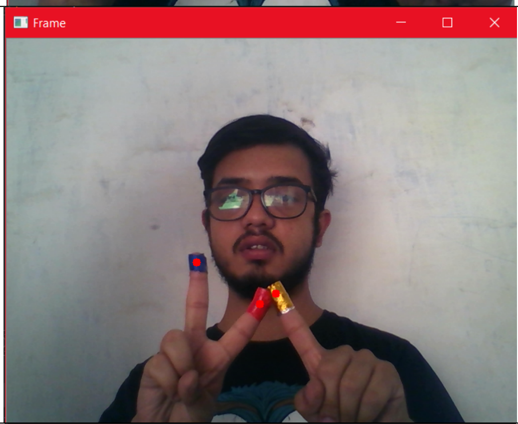
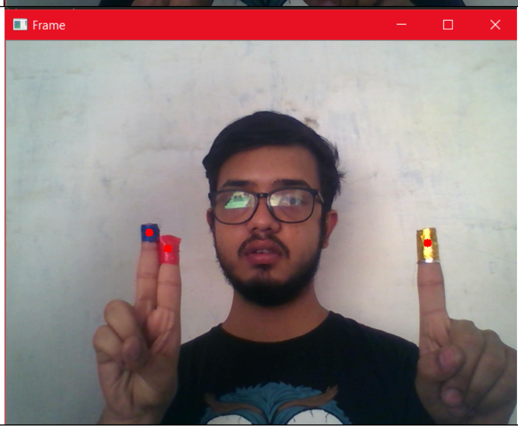
Gesture	Action Performed
	Free Movement
	Left Click
	Right Click

Table 4: Mapping for Virtual Mouse Control

Chapter 5

Summary and Future Work

5.1 Summary

In this project, the main aim was to implement a gesture controlled system capable of detecting a gesture being performed as well as perceive what it means and respond accordingly. It basically involves study and application of various concepts of Image Processing for the achievement of the aforesaid goal. Image Processing here has been applied for implementation of Virtual mouse and Virtual keyboard which will be the next step in the evolution of the compact electronics. Background subtraction algorithm has been used for virtual keyboard control while colour detection algorithm has been used for virtual mouse control.

5.2 Future Work

- The current solution involves use of algorithms which have some limitations like three different colours are required to detect three different points. In future we will want to develop an algorithm which does not require any extra stuff

- In virtual keyboard control the camera has to be kept fixed, this problem will be solved by developing a better algorithm.

- The number of permutations of gestures which can be uniquely identified with the given algorithm is also limited. The algorithm can be extended to involve more number of gestures.
- Also, the current solution can be integrated with multiple existing technologies like IOT, Web Applications, Mobile Applications to increase the domain of application.

References

[1] Shaunak Chokshi, Shashank Sharma, Hardik Joshi, Gesture Controlled Robot, International Journal of Scientific Engineering Research, Volume 7, Issue 7, July-2016

[2] Virtual Keyboard Using Image Processing, Int. Journal of Engineering Research and Applications ISSN : 2248-9622, Vol. 4, Issue 4(Version 4), April 2014

[3] Ankur Yadav, Anand Pandey, Avinash Singh, Ajit Kr Singh Computer Mouse Implementation using Object Detection and Image Processing International Journal of Computer Applications (0975 8887) Volume 69 No.21, May 2013

[4] S. Kolkur, D. Kalbande, P. Shimpi, C. Bapat, and J. Jatakia, Human Skin Detection Using RGB, HSV and YCbCr Color Models, ICCASP/ICMMD-2016, Advances in Intelligent Systems Research. Vol. 137, Pp. 324-332.

[5] Asanterabi Malima, Erol Ozgur, and Mujdat Cetin, "A Fast Algorithm for Vision-Based Hand Gesture Recognition for Robot Control", IEEE International Conference on Computer Vision, 2006.

[6] Image Processing tutorials by TutorialsPoint https://www.tutorialspoint.com/dip/index.html

[7] Image Processing: https://en.wikipedia.org/wiki/Image%25processing