A PROJECT REPORT

ON

“COMPUTER VISION: GESTURE RECOGINITION, USER INTERACTION, CONTROLS AND MANIPULATION”

Bachelor of Technology

Computer Science and Engineering

BY

Alfran Ali 2015KUCP1032

Shams Ali 2015KUCP1034

Jai Pal Singh 2015KUCP1038

UNDER THE GUIDANCE OF

DR. NEERAJ KUMAR RAO

DR. SMITA NAVAL



DEPARTMENT OF COMPUTER SCIENCE AND ENGINEERING

Indian Institute of Information Technology, Kota(India)

2015-2019

Declaration

I hereby declare that the work reported in the B.tech report entitled Computer Vision: Gesture Recognition, User Interaction, Control and Manipulation submitted at Indian Institute of Information Technology, Kota is an authentic record of my work carried under the supervision of Dr. Neeraj Kumar Rao and Dr. Smita Naval. I have not submitted this work elsewhere for any other degree.

............

ALFRAN ALI

(2015KUCP1032)

SHAMS ALI

(2015KUCP1034)

JAI PAL SINGH

(2015KUCP1038)

Under the supervision of

Dr. NEERAJ KUMAR RAO

Dr. SMITA NAVAL

Department of Computer Science and Engineering

Indian Institute of Information Technology, Kota

Acknowledgements

We are profoundly grateful to Dr. Neeraj Kumar Rao and Dr. Smita Naval for their expert guidance and continuous encouragement throughout to see that this project rights its target since its commencement. Their timely and efficient contribution helped us shape our work into its final form and we express our sincerest gratitude for their assistance in any way that we may have asked. We appreciate their guidance in our project that has improved our project many folds, thanks for the comments and advise.

We would also like to thank Dr. Amit Kumar and all other faculty members of IIIT Kota for their direct or indirect supports and advise without which this project would not have been possible.

ALFRAN ALI

SHAMS ALI

JAI PAL SINGH

ABSTRACT

As technology has become one of the most important part of the day to day human life, so in order to make the best use of those technologies the study of the relationship and interaction between human and computer i.e. called human computer interaction(HCI) is important. In this way with the help of human computer interaction we can develop and improve computer system to serve the needs of human in a better and efficient way. Human computer interaction can be applied in various fields and research areas as like in medical system which would be beneficial for the elderly people who are not able to walk or talk or express their feelings by words. It is also helpful in the field of computer application where we can make the use of human computer interaction with the help hand gesture recognition to control the media player operations as like play or pause.

In this project we are trying to provide a non-tangible way of human computer interaction and that is with the help of hand gesture recognition to control different operations like play or pause of media player. For the recognition of the hand gesture we are using Convex-Hull algorithm which try to find different peak points of the hand. After the recognition of the hand gesture we will bind it with different operation of the media player accordingly.

Keywords: Human Computer Interaction, Hand Gesture Recognition, Convex-hull algorithm, Media Player

Contents

[Introduction 1](#_Toc513143733)

[1.1 Computer Vision 1](#_Toc513143734)

[1.2 Gesture Recognition 1](#_Toc513143735)

[1.3 Human Computer Interaction 2](#_Toc513143736)

[1.4 Hand Gesture Recognition 2](#_Toc513143737)

[1.5 Importance of Human Computer Interaction 2](#_Toc513143738)

[Motivation and Objective 4](#_Toc513143739)

[2.1 Motivation 4](#_Toc513143740)

[2.2 Objective 4](#_Toc513143741)

[Literature Survey 5](#_Toc513143742)

[3.1 Look Based Media Player 5](#_Toc513143743)

[3.2 MP-FEG 6](#_Toc513143744)

[3.3 Gesture Recognition and fingertip detection for HCI 6](#_Toc513143745)

[3.4 Intuitive HDI for Video Control and Feedbac 7](#_Toc513143746)

[3.5 Vision-Based Hand Gesture Recognition Using Blob Analysis 8](#_Toc513143747)

[Methodology 9](#_Toc513143748)

[Result and Discussions 13](#_Toc513143749)

[5.1 Result 13](#_Toc513143750)

[5.2 Discussion 13](#_Toc513143751)

[Future Work 14](#_Toc513143752)

[References 15](#_Toc513143753)

List of Figures

[Figure 1 - Number of publications v/s Year (Nardi, 2016) 3](#_Toc513143754)

[Figure 2 - Block diagram of proposed method 9](#_Toc513143755)

[Figure 3 - Convex Hull (Hand Gesture Recognition) 10](#_Toc513143756)

[Figure 4 - Gesture Recognition 12](#_Toc513143757)

Chapter 1

# Introduction

In the present world, the interaction with the computing devices has advanced to such an extent that as humans it has become necessity and we cannot live without it. The technology has become so embedded into our daily lives that we use it to work, shop, communicate and even entertain our self. It has been widely believed that the computing, communication and display technologies progress further. To efficiently use them, most computer applications require more and more interaction. For that reason, human-computer interaction (HCI) has been a lively field of research in the last few years. Recognizing hand gestures for interaction can help in achieving the ease and naturalness desired for human computer interaction.

## Computer Vision

Computer vision is a field of computer science that is entrusted with how computers can gain high level understanding from digital images or sequences of digital images or videos. In other words, we can say that it provides computers the ability to analyse and understand useful information from digital images or sequence of images or videos. Actually, computer vision gives computers the ability to see or visualise things in the same way as we see or visualise things through human vision. Computer Vision provides us with an opportunity to make the best and efficient use of the computer technologies and improve the computer system to serve the various human needs.

## 1.2 Gesture Recognition

Gesture recognition is the branch of computer science that deals with the mathematical interpretations of the human motion with the help of a computing device. The origin of the gesture can from any part of the body but most commonly its origin is either face or hand. In simple words we can say that gesture recognition is used to recognise and identify the movements of the different parts of the human body. The most common gesture recognitions are hand gesture recognition, face recognition, facial expression recognition etc. Gesture recognition is one of the best way of interaction with the computer system.

## Human Computer Interaction

Human computer interaction is the branch of computer science that deals with the relationship and interaction of the human with computer system. In simple words we can say that it enables computers to interact with the humans in the similar way as humans interact with humans. Human computer interaction can be done in many ways as like through hand gesture recognition, voice recognition, speech recognition, face recognition, facial expression recognition etc. In fact, we can say that human computer interaction is laying the foundation of a future where humans and computers would be related to or linked with each other as they were never before.

* 1. Hand Gesture Recognition

Hand gesture recognition is considered as one of the best interaction style and non-tangible technique for human computer interaction. In this technique different hand gestures are recognised using some algorithm like convex-hull etc. and certain actions are being performed based on those gestures. Hand gesture can be categorised into two types of hand gesture representation:

a) Contact Based: In contact-based hand gesture representation, the hand must be in contact with some of the external device as like gloves for the gesture to be recognised. Without the external device contact-based hand gesture representation is meaningless.

b) Vision Based: In vision-based hand gesture representation the hand need not to be in contact with some external device to recognise the gesture. The only requirement for this type of hand gesture representation is that you need a computer inbuilt with camera or computer system with camera.

* 1. Importance of Human Computer Interaction

As we all know that human computer interaction deals with how humans and computers interact and how they are related to each other. So, the study of human computer interaction is important because if we want to improve the computer system to serve the human need we must know about how they are related with each other also how they interact with each other. Human computer interaction has gained its importance and has become one of the most interesting domain for the researchers to work on such related topics. Over the past few decades the number of publications on the topic related with human computer interaction has increased exponentially and also it is increasing day by day.

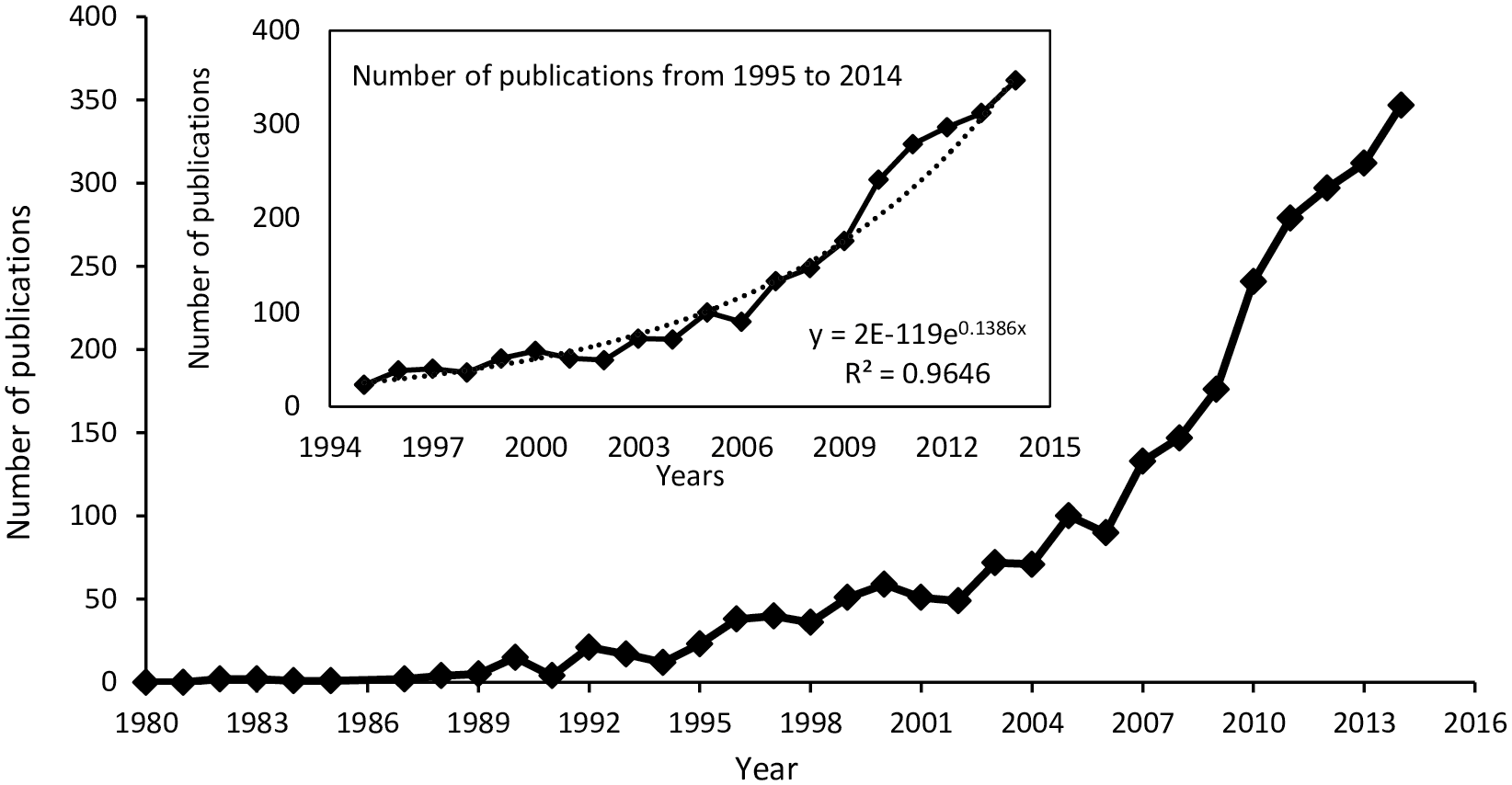


Figure - Number of publications v/s Year (Nardi, 2016)

Chapter 2

# Motivation and Objective

## 2.1 Motivation

The motivation behind this project is as follows:

1. To provide and improve the existing non-tangible (without being in contact) way of human computer interaction using hand gesture recognition. So, in this project we are using vision-based hand gesture representation technique to improve the computer to serve the human need.
2. Vision based hand gesture representation is helpful for elderly people as well as for those people who are not able to speak or express their feelings through words because we can use those gestures and bind some action related to those gestures so that they can be performed when recognised without any difficulty.
3. This area of research is useful for various other research areas as like it would be useful in the field of medical science, speech recognition, and other research areas where human body motion and gestures are combinedly being analysed.

## Objective

The goal of our project is to control video player with the help of hand gestures. We have set the following objectives for our video player to achieve the goal:

1. In order to achieve the above stated objective first we need to develop a model that can recognise different hand gestures according to the need.
2. And in order to recognise different hand gestures first we need to create and define hand gestures that we want to use in our project.
3. Then we need to implement that model using appropriate algorithms that can increase performance and efficiency of the video player.

Chapter 3

# Literature Survey

For the approaches used in developing algorithm of the hand gesture recognition and user interaction system, there are various techniques implementing for detecting and tracking the dynamic hand gestures such as Haar-like features, convex hull, contour matching, and skin colour. Here is the literature we went through to gain knowledge about this field.

## 3.1 Look Based Media Player (Tripath, Thakur, Inchanale, & Sanghavi, 2017)

The objective of this project is to develop an advanced media player which plays and pauses the video by detecting whether the user is looking at the screen or not using a web camera. If yes then doesn't interrupts the video and allows it to play. In case if the user is not looking at the or say the system couldn't detect the users face then it immediately stops the video. Also, they want toadd a feature of controlling other features of media player such as volume up and volume down, backward and forward video using hand gestures.

**Algorithm:**

The objective is achieved by using Haar cascade classifier. Haar cascade is an object detection algorithm, mainly used to locate faces. In this, the cascade function is trained from a lot of positive and negative images. It is then used to detect objects in other images.

**Advantage:**

The main of cascading is achieving a good performance for both accuracy and time-complexity, when you have a number of weak classifier.

**Disadvantage:**

The problem associated with this implementation is that it requires very high computation.

## 3.2 MP-FEG: Media Player controlled by Facial Expressions and Gestures (Agarwal, 2015)

The objective is to design a Media Player system controller by Facial  
Expressions and Gestures (MP-FEG) by using Support Vector Machine(SVM). This technique is a non-tangible way to interact with computer.

**Algorithm:**

The objective is achieved by detecting and tracking one landmark point on the finger and 18 landmark points on the lips to capture the movement of the finger and the lips of the user. The movement patterns are classified into hand gestures and facial expressions using support vector machine (SVM) and Feature Point Tracking.

**Advantage:**

By this method, authors have achieved *∼*98.65% and *∼*100% recognition accuracies for hand-gestures and facial expressions respectively. Occurrence of each of theseactions (5 hand-gestures and 3 facial expressions) is associated with a command to control (e.g., to select, play, pause the video) the video player.

**Disadvantage:**

The problem associated with this implementation is that it takes around 5 seconds to recognize 2 expressions or gestures, which is a very large time period.

## 3.3 Gesture Recognition and fingertip detection for HCI (Gunasundari, Prakash, Deepa, & Kasthuri, 2017)

The paper proposes a novel gesture recognition and fingertip detection algorithm for Human Computer Interaction in particular mouse control operations using real time camera. The hand gestures are captured using real time camera**.**

**Algorithm:**

Authors have used region growing algorithm followed by morphological operations to segment hand region. The centroid of the palm region is calculated and the finger tips are then detected using the convex hull algorithm.

The proposed method is tested on five different gestures and the results prove that the gestures are able to be recognized and the finger tips detected. The method can be applied for hand gesture-controlled mouse operations**.**

**Advantage:**

1. Region growing methods can correctly separate the regions that have the same properties we define.
2. It can provide the original images which have clear edges with good segmentation results.

**Disadvantage:**

This algorithm is computationally expensive and is sensitive to noise.

## 3.4 Intuitive Human-Device Interaction for Video Control and Feedback (Pessemier, Martens, & Joseph, 2017)

The paper proposes to use commercially available motion sensing input devices for facial recognition, gesture control, and speech recognition to automatically authenticate a user, enable video control (play, pause, seeking), and browsing, selecting, and rating content by hand gestures or voice commands. By monitoring the user’s gaze and using emotion recognition techniques, the user’s interests and engagement with the content can be estimated

**Algorithm:**

The authors have used K Nearest Neighbour(KNN) and Support Vector Machine(SVM) algorithms to achieve their goal.

**Advantage:**

The algorithm isautomatically non-linear. It can detect linear or non-linear distributed data. It tends to perform very well with a lot of data points.

**Disadvantage:**

The major drawback of this method is that the other peripheral devices are required to achieve the goal. Such devices may not be owned by everyone.

## 3.5 The Vision-Based Hand Gesture Recognition Using Blob Analysis (Ganokratanaa & Pumrin, 2017)

The paper proposes that human-computer interaction (HCI) can be applied in various areas including medical system and the development of algorithm by using hand gestures**.**

**Algorithm**

This paper proposes a dynamic hand gesture recognition algorithm for elder people. The algorithm implements in a vision-based hand gesture recognition using optical flow and blob analysis to track six dynamic hand gestures and classify their meanings.

**Advantage:**

The experiment provided good results for all six hand gestures in detection, tracking, and classification procedures. Also, no other external device is required except web camera which is generally available to all.

This technique includes high flexibility and excellent performance.

**Disadvantage:**

This technique requires clear background-foreground relation and high pixel-precision.

Chapter 4

# Methodology

The block diagram of the proposed method is shown in Figure 2. The video of the hand gesture is captured using the open cv tool. Video frames are stored as NumPy arrays.

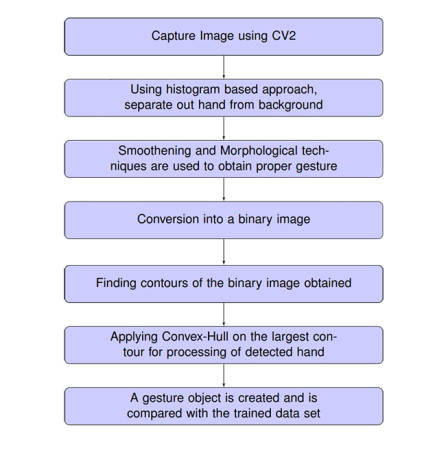


Figure - Block diagram of proposed method

First the program captures the background. This will help reduce detection of false stationary objects which are around the same histogram range as our hand. Now that we have a plain background image detected, we go for capturing a histogram of our hand that we need the program to recognize in each frame. For this the user moves his hand to cover the 9 small boxes in such a way that all or most of his hand shades are covered in them with no air gaps or dark shadows on it. This step is very necessary for a proper recognition. Then convert image frame (with hand on the 9 small boxes) to HSV colour space. Extract the pixel values from those 9 boxes in HSV colour space and create and return a histogram using those pixel values.

After generating histogram, apply morphology and smoothening techniques (Bilateral and Median blur) to the back projection generated. Then apply threshold to generate a binary image from the back projection. This threshold is used as a mask to separate out the hand from the rest of the frame. The next step is to find out contours of the image generated in the previous step. We first find out all contours of the image and then validate the largest contour to verify if it matches the profile of a hand or not. Now that we have the largest contour, the program then works only on this largest contour found. We find the convex hull of the contour.

Once we have the convex hull of the largest contour, we go for detecting the centre of the hand. We find the range of coordinates that span the largest contour (extreme end points that form a rectangle that bounds the contour). In the range determined thus, run a loop that takes every point in the rectangular range and measures the distance from that point to the nearest point on the contour. We find the point with the largest such distance. The point is the centre of the largest circle that can be inscribed inside the contour and the distance is the radius of the circle. We check if the hand radius is large enough because very small radius might suggest falsely detected objects in case the frame is actually kept empty.

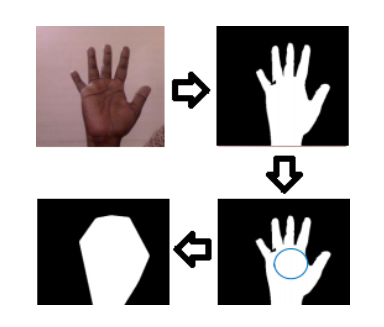


Figure - Convex Hull (Hand Gesture Recognition)

As shown in figure 3, first we have input frame of hand gesture, then we have segmented hand region from the background by applying morphology and smoothening techniques. Next step in the figure is detecting the centre of the hand. The image shows segmented hand region with plotted radius. Last image is the convex hull of the hand region.

Once we have the palm centre coordinates and radius, we go for finding fingers in our hands. After this we have palm centre, radius and exactly only those many points in a convex hull that belong to our fingers. From these parameters, we can then figure out the angles between different fingers, the count of fingers shown and ultimately the gesture defined with those parameters.

From the figure 4, it is very clear that from the above method count of finger tips is detected. It does not depend on the shape of the hand gesture. It is only concerned with the finger tips including the thumb. This method works perfectly for both front and back side of hand. The image (a) shows four fingers and one thumb gesture. Here total count is five. Our method typically works on determining the count only.

In figure 4(b) and 4(c) the count of the fingers is 3. Although the gesture is different but as the method is only concerned only with finger tip counts, so they are not different for our method.

In figure 4(d), the finger count is zero as the maximum length of finger tip contours is approximately equal to the radius of the circle at the centre of hand. Thus, the method considers this as the zero-finger gesture.



Figure - Gesture Recognition

Chapter 5

# Result and Discussions

## 5.1 Result

Following is the outcome of our applied methodology:

1. Successfully recognises different gestures of hand.
2. Works for both front side and back side of the hand.
3. Does not work for those gestures in which finger tips are pointing downwards.
4. It will work better if the background colour is distinct from the colour of the hand.
5. For the gesture to be recognised, the hand should be in the front of top-right portion of the screen.
6. Gesture would not be recognised if the background is too shiny.
7. Too much movement in the background will affect the gesture recognition process.
8. For now, we use different keys for different instant of the gesture recognition.

## 5.2 Discussion

1. First our method takes image of the background without placing the hand in front of the web camera.
2. Then after placing hand, it takes the images of hand from the video frames
3. From that image it tries to remove the background of the image captured.
4. After that bilateral and median filters are applied for smoothening of the captured image.
5. The corresponding gesture is recognised.

Chapter 6

# Future Work

Following is the work that we intent to complete in the future:

1. In future we want to extend this work and try to use it to control different operations of the media player with the help of the gestures recognised. The operations we want to control includes
   1. Play/Pause a video
   2. Stop a video
   3. Play next video of the playlist
   4. Play previous video of the playlist
2. Also, to improve the non-tangible way of human computer interaction and improving the computer system to serve the human need.

Chapter 7

# References

Agarwal, S. &. (2015). MP-FEG: Media Player controlled by Facial. *2015 Fifth National Conference on Computer Vision, Pattern Recognition, Image Processing and Graphics (NCVPRIPG)*, (pp. 1-4).

Ganokratanaa, T., & Pumrin, S. ( 2017). The Vision-Based Hand Gesture Recognition Using. *2017 International Conference on Digital Arts, Media and Technology (ICDAMT)*, 336-341.

Gunasundari, T., Prakash, R. M., Deepa, T., & Kasthuri, N. (2017). Gesture Recognition and finger tip detection for. *Innovations in Information, Embedded and Communication Systems (ICIIECS)* (pp. 1-4). International Conference.

Nardi, P. (2016). Structure and Evolution of Mediterranean Forest Research: A Science Mapping Approach. *Pone*, 8.

Pessemier, T. D., Martens, L., & Joseph, W. (2017). *Intuitive Human-Device Interaction for Video.*

Tripath, R., Thakur, R., Inchanale, S., & Sanghavi, N. (2017). Look Based Media Player. *International Research Journal of Engineering and Technology (IRJET)*, 5.